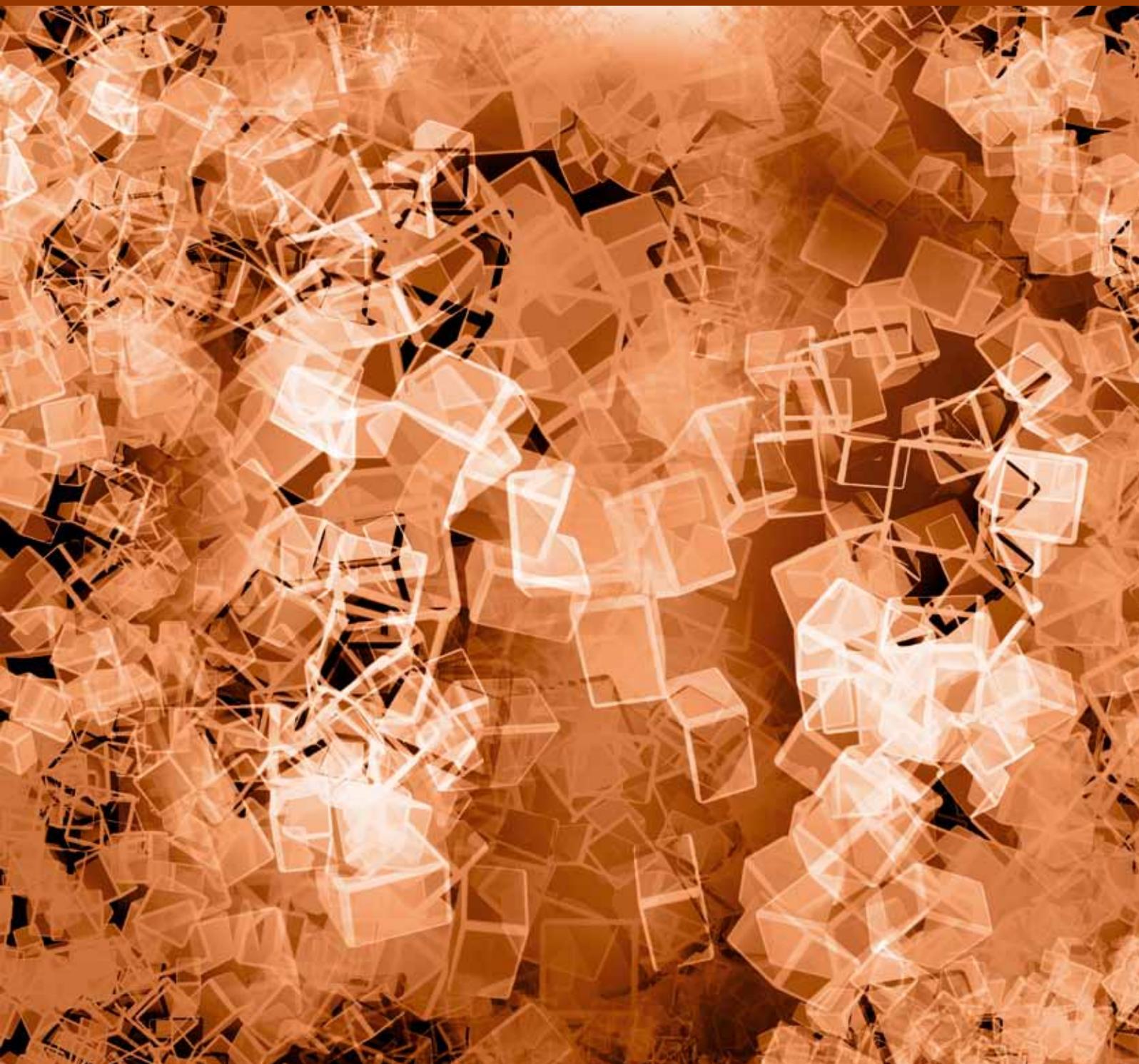


DISCRETE DYNAMICS IN NATURE AND SOCIETY

# GREEN INTELLIGENT TRANSPORT SYSTEM

GUEST EDITORS: WUHONG WANG, GEERT WETS, HEINER BUBB, AND HUIMIN NIU





---

# **Green Intelligent Transport System**

Discrete Dynamics in Nature and Society

---

## **Green Intelligent Transport System**

Guest Editors: Wuhong Wang, Geert Wets, Heiner Bubb,  
and Huimin Niu



---

Copyright © 2014 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "Discrete Dynamics in Nature and Society." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Editorial Board

Mustapha A. Rami, Spain  
Douglas R. Anderson, USA  
Viktor Avrutin, Germany  
Stefan Balint, Romania  
Kenneth S. Berenhaut, USA  
Gabriele Bonanno, Italy  
Elena Braverman, Canada  
Pasquale Candito, Italy  
Jinde Cao, China  
Wei-Der Chang, Taiwan  
Cengiz Çinar, Turkey  
Daniel Czamanski, Israel  
Richard H. Day, USA  
Manuel De la Sen, Spain  
Josef Diblík, Czech Republic  
Xiaohua Ding, China  
Adam Doliwa, Poland  
Zengji Du, China  
Elmetwally Elabbasy, Egypt  
Hassan A. El-Morshedy, Egypt  
Daniele Fournier-Prunaret, France  
Vladimir Gontar, Israel  
Luca Guerrini, Italy  
Taher S. Hassan, Egypt  
Xue Z. He, Australia  
Weihong Huang, Singapore  
Bratislav Iricanin, Serbia  
Giuseppe Izzo, Italy  
Lee C. Jang, Korea  
Jun Ji, USA  
Zhen Jin, China  
Nikos I. Karachalios, Greece  
Eric R. Kaufmann, USA

Recai Kilic, Turkey  
Ryusuke Kon, Japan  
Victor S. Kozyakin, Russia  
Mustafa Kulenović, USA  
Jürgen Kurths, Germany  
Wei Nian Li, China  
Xiaojie Lin, China  
Xiaohui Liu, UK  
Ricardo López-Ruiz, Spain  
Qing-hua Ma, China  
Akio Matsumoto, Japan  
Rigoberto Medina, Chile  
Eleonora Messina, Italy  
Ugurhan Mugan, Turkey  
Pham H. A. Ngoc, Vietnam  
Piyapong Niamsup, Thailand  
Juan J. Nieto, Spain  
Prasanta K. Panigrahi, India  
Garyfalos Papashinopoulos, Greece  
Beatrice Paternoster, Italy  
Mingshu Peng, China  
Carlo Piccardi, Italy  
Peter Plath, Germany  
Chuanxi Qian, USA  
Vladimir S. Rabinovich, Mexico  
Irena Rachůnková, Czech Republic  
Morteza Rafei, The Netherlands  
Aura Reggiani, Italy  
Pavel Rehak, Czech Republic  
Marko Robnik, Slovenia  
Yuriy Rogovchenko, Norway  
Ventsi G. Rumchev, Australia  
Budhi S. D. Sagar, India

R. Sahadevan, India  
Jagannathan Sarangapani, USA  
Leonid Shaikhet, Ukraine  
Vimal Singh, Turkey  
Seenith Sivasundaram, USA  
Francisco Solis, Mexico  
Gualberto Solís-Perales, Mexico  
Yihong Song, China  
Michael Sonis, Israel  
Jian-Ping Sun, China  
Stepan Tersian, Bulgaria  
Gerald Teschl, Austria  
Tetsuji Tokihiro, Japan  
Delfim F. M. Torres, Portugal  
Firdaus Udwardia, USA  
Antonia Vecchio, Italy  
Qi-Ru Wang, China  
Gongnan Xie, China  
Ibrahim Yalcinkaya, Turkey  
Xiang P. Yan, China  
Xiaofan Yang, China  
Bo Yang, USA  
Her-Terng Yau, Taiwan  
Jianshe Yu, China  
Agacik Zafer, Turkey  
Guang Zhang, China  
Binggen Zhang, China  
Zhengqiu Zhang, China  
Lu Zhen, China  
Baodong Zheng, China  
Yong Zhou, China  
Zhan Zhou, China  
Zuo-nong Zhu, China

## Contents

**Green Intelligent Transport System**, Wuhong Wang, Geert Wets, Heiner Bubb, and Huimin Niu  
Volume 2014, Article ID 250475, 2 pages

**Passenger Flow Prediction of Subway Transfer Stations Based on Nonparametric Regression Model**, Yujuan Sun, Guanghou Zhang, and Huanhuan Yin  
Volume 2014, Article ID 397154, 8 pages

**Activity-Trip Chaining Behavior of Urban Low-Income Populations in Nanjing, China: A Structural Equations Analysis**, Zhaoming Chu, Hui Chen, Lin Cheng, Xuewu Chen, and Senlai Zhu  
Volume 2014, Article ID 360269, 11 pages

**Rough Set Approach for Group Evacuation Behavior Analysis in Passenger Transport Hub Area**, Peng Chen, Xiaofeng Ma, and Shunying Zhu  
Volume 2014, Article ID 682147, 9 pages

**A Bayesian Network Model for Origin-Destination Matrices Estimation Using Prior and Some Observed Link Flows**, Lin Cheng, Senlai Zhu, Zhaoming Chu, and Jingxu Cheng  
Volume 2014, Article ID 192470, 9 pages

**A Conceptual Architecture for Adaptive Human-Computer Interface of a PT Operation Platform Based on Context-Awareness**, Qing Xue, Xuan Han, Mingrui Li, and Minxia Liu  
Volume 2014, Article ID 371204, 7 pages

**Mixture Augmented Lagrange Multiplier Method for Tensor Recovery and Its Applications**, Huachun Tan, Bin Cheng, Jianshuai Feng, Li Liu, and Wuhong Wang  
Volume 2014, Article ID 914963, 9 pages

**Modeling the Joint Choice Decisions on Urban Shopping Destination and Travel-to-Shop Mode: A Comparative Study of Different Structures**, Chuan Ding, Binglei Xie, Yaowu Wang, and Yaoyu Lin  
Volume 2014, Article ID 492307, 10 pages

**Effectiveness of Variable Message Signs on Driving Behavior Based on a Driving Simulation Experiment**, Xuedong Yan and Jiawei Wu  
Volume 2014, Article ID 206805, 9 pages

**Effect of Height on Pedestrian Route Choice between Stairs and Escalator**, Qian Li, Changxu Ji, Limin Jia, and Yong Qin  
Volume 2014, Article ID 965305, 6 pages

**Optimizing Crew Rostering with Multilicense on High-Speed Railway Lines**, Zhiqiang Tian  
Volume 2014, Article ID 910569, 10 pages

**Modeling the Perceptions and Preferences of Pedestrians on Crossing Facilities**, Hongwei Guo, Facheng Zhao, Wuhong Wang, Yanlong Zhou, Yujie Zhang, and Geert Wets  
Volume 2014, Article ID 949475, 8 pages

**Application of the Expanded Theory of Planned Behavior in Intercity Travel Behavior**, Jing Peng, Juan Zhi-cai, and Gao Lin-jie  
Volume 2014, Article ID 308674, 10 pages

**Conspicuity Research on the Highway Roadside Objects: A Simulator Study**, Xuemei Chen, Zhonghua Wei, Xia Zhao, Mingyang Hao, and Tongyang Zhang  
Volume 2014, Article ID 864791, 9 pages

**A Quasi-Poisson Approach on Modeling Accident Hazard Index for Urban Road Segments**, Lu Ma, Xuedong Yan, and Wenxin Qiao  
Volume 2014, Article ID 489052, 8 pages

**Pedestrian Detection and Tracking for Counting Applications in Metro Station**, Chen Yan-yan, Chen Ning, Zhou Yu-yang, Wu Ke-han, and Zhang Wei-wei  
Volume 2014, Article ID 712041, 11 pages

**Pedestrian Guiding Signs Optimization for Airport Terminal**, Jianxin Lin, Rui Song, Jifeng Dai, and Pengpeng Jiao  
Volume 2014, Article ID 125910, 14 pages

**A Bilevel Programming Model to Optimize Train Operation Based on Satisfaction for an Intercity Rail Line**, Zhipeng Huang and Huimin Niu  
Volume 2014, Article ID 432096, 7 pages

**Capacity Estimation for On-Ramp Merging Section of Urban Expressway Based on Time Headway Loss**, Xing-jian Xue, Feng Shi, and Qun Chen  
Volume 2014, Article ID 107601, 9 pages

**Research and Application of the Beijing Road Traffic Prediction System**, Ruimin Li, Hongliang Ma, Huapu Lu, and Min Guo  
Volume 2014, Article ID 316032, 8 pages

**Dynamic Network Design Problem under Demand Uncertainty: An Adjustable Robust Optimization Approach**, Hua Sun, Ziyou Gao, and Fangxia Zhao  
Volume 2014, Article ID 739656, 16 pages

**Congestion Behavior under Uncertainty on Morning Commute with Preferred Arrival Time Interval**, LingLing Xiao, Ronghui Liu, and HaiJun Huang  
Volume 2014, Article ID 767851, 9 pages

**Modeling the Equilibrium Bus Line Choice Behavior and Transit System Design with Oblivious Users**, Chuan-Lin Zhao and Hai-Jun Huang  
Volume 2014, Article ID 173876, 5 pages

**Statistical Analysis of Traffic Accidents in Shanghai River Crossing Tunnels and Safety Countermeasures**, Linjun Lu, Jian Lu, Yingying Xing, Chen Wang, and Fuquan Pan  
Volume 2014, Article ID 824360, 7 pages

**Automated Generation of Traffic Incident Response Plan Based on Case-Based Reasoning and Bayesian Theory**, Yongfeng Ma, Wenbo Zhang, Jian Lu, and Li Yuan  
Volume 2014, Article ID 920301, 7 pages

**Optimization Based High-Speed Railway Train Rescheduling with Speed Restriction**, Li Wang, Wenting Mo, Yong Qin, Fei Dou, and Limin Jia  
Volume 2014, Article ID 934369, 14 pages

**An Operation Optimization for Express Freight Trains Based on Shipper Demands**, Yuzhao Zhang and Yusong Yan  
Volume 2014, Article ID 232890, 8 pages

**Development of a Safety Evaluation Model for Provincial Highway**, Li Yuan, He-wei Yuan, Yong-feng Ma, and Ying-wei Ren  
Volume 2014, Article ID 394846, 10 pages

**Tour Route Multiobjective Optimization Design Based on the Tourist Satisfaction**, Yan Han, Hongzhi Guan, and Jiaying Duan  
Volume 2014, Article ID 603494, 8 pages

**The Research of the Driver Attention Field Modeling**, Pengfei Tao, Hongyu Hu, Zhenhai Gao, Xin Liu, Xianmin Song, Yan Xing, Yuzhou Duan, and Fulu Wei  
Volume 2014, Article ID 270616, 9 pages

**A Study on High-Speed Rail Pricing Strategy in the Context of Modes Competition**, Enjian Yao, Qirong Yang, Yongsheng Zhang, and Xun Sun  
Volume 2013, Article ID 715256, 6 pages

**Modeling of a Small Transportation Company's Start-Up with Limited Data during Economic Recession**, Xiaoping Fang, Jake Ansell, and Weiya Chen  
Volume 2013, Article ID 802528, 10 pages

**Multinomial Logit Model of Pedestrian Crossing Behaviors at Signalized Intersections**, Zhu-Ping Zhou, Ying-Shun Liu, Wei Wang, and Yong Zhang  
Volume 2013, Article ID 172726, 8 pages

**Pattern Analysis of Driver's "Pressure-State-Response" in Traffic Congestion**, Weiwei Qi, Yulong Pei, Mo Song, and Yiming Bie  
Volume 2013, Article ID 853845, 11 pages

**The Effects of the Tractor and Semitrailer Routing Problem on Mitigation of Carbon Dioxide Emissions**, Hongqi Li, Yanran Li, Yue Lu, Qiang Song, and Jun Zhang  
Volume 2013, Article ID 809135, 14 pages

**Comparison of Electric Vehicle's Energy Consumption Factors for Different Road Types**, Enjian Yao, Zhiqiang Yang, Yuanyuan Song, and Ting Zuo  
Volume 2013, Article ID 328757, 7 pages

**Efficient Processing of Continuous Skyline Query over Smarter Traffic Data Stream for Cloud Computing**, Wang Hanning, Xu Weixiang, Jiulin Yang, Lili Wei, and Jia Chaolong  
Volume 2013, Article ID 209672, 10 pages

**Transit Station Congestion Index Research Based on Pedestrian Simulation and Gray Clustering Evaluation**, Shu-wei Wang, Li-shan Sun, Jian Rong, and Zi-fan Yang  
Volume 2013, Article ID 891048, 8 pages

**Performance of the Priority Control Strategies for Bus Rapid Transit: Comparative Study from Scenario Microsimulation Using VISSIM**, Min Yang, Wei Wang, Bo Wang, and Jing Han  
Volume 2013, Article ID 398938, 9 pages

**The Analysis of Braess' Paradox and Robustness Based on Dynamic Traffic Assignment Models**, Bai-Bai Fu, Chun-Xue Zhao, and Shu-bin Li  
Volume 2013, Article ID 796842, 8 pages

**Development of Degree-of-Priority Based Control Strategy for Emergency Vehicle Preemption Operation**, Jiawen Wang, Wanjing Ma, and Xiaoguang Yang  
Volume 2013, Article ID 283207, 10 pages

**Optimal Design of the Feeder-Bus Network Based on the Transfer System**, Lianbo Deng, Wei Gao, Yanbing Fu, and Wenliang Zhou  
Volume 2013, Article ID 483682, 10 pages

**Study on Driving Stability of Tank Trucks Based on Equivalent Trammel Pendulum for Liquid Sloshing,**

Xian-sheng Li, Xue-lian Zheng, Yuan-yuan Ren, Yu-ning Wang, and Zhu-qing Cheng

Volume 2013, Article ID 659873, 15 pages

**Dynamic Analysis of Traffic State and Congestion Propagation on Bidirectional Grid Network,**

Shu-bin Li, Bai-bai Fu, and Jian-feng Zheng

Volume 2013, Article ID 165086, 7 pages

**A Hybrid Temporal-Spatio Forecasting Approach for Passenger Flow Status in Chinese High-Speed Railway Transport Hub,** Zhengyu Xie, Limin Jia, Yong Qin, and Li Wang

Volume 2013, Article ID 239039, 7 pages

**Exploring the Effects of Different Walking Strategies on Bi-Directional Pedestrian Flow,** Lili Lu,

Gang Ren, Wei Wang, Chen Yu, and Chenzi Ding

Volume 2013, Article ID 150513, 9 pages

**Optimal Skip-Stop Schedule under Mixed Traffic Conditions for Minimizing Travel Time of Passengers,**

Sun Feng, Zhu Wen-tao, Ye Ying, and Wang Dian-hai

Volume 2013, Article ID 930562, 8 pages

**Using CVIS to Improve Bus Schedule Adherence: A Predictive Control Strategy and Its**

**Hardware-in-the-Loop Field Tests,** Wei Yin, Jing Teng, Xiaoguang Yang, and H. Michael Zhang

Volume 2013, Article ID 171582, 8 pages

**Robustness Measure of China's Railway Network Topology Using Relative Entropy,**

Fenling Feng and Lei Wang

Volume 2013, Article ID 391709, 8 pages

**A Review Study on Traction Energy Saving of Rail Transport,** Xuesong Feng, Hanxiao Zhang, Yong Ding, Zhili Liu, Hongqin Peng, and Bin Xu

Volume 2013, Article ID 156548, 9 pages

## Editorial

# Green Intelligent Transport System

Wuhong Wang,<sup>1</sup> Geert Wets,<sup>2</sup> Heiner Bubb,<sup>3</sup> and Huimin Niu<sup>4</sup>

<sup>1</sup> Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

<sup>2</sup> Transportation Research Institute (IMOB), Hasselt University, 3590 Diepenbeek, Belgium

<sup>3</sup> Lehrstuhl für Ergonomie, Technische Universität München, 85747 Munich, Germany

<sup>4</sup> School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

Correspondence should be addressed to Wuhong Wang; wangwuhong@bit.edu.cn

Received 29 April 2014; Accepted 29 April 2014; Published 14 May 2014

Copyright © 2014 Wuhong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The transportation system is a complex dynamic system with various elements such as drivers, passengers, pedestrians, vehicles, and roads. Recently, with the rapid development of motorization, transportation problems (i.e., accidents, congestions, and pollutions) have attracted much attention from the public. Not only are these problems highly relevant to our daily life, but also they provide interesting phenomena and talking-points. For example, traffic congestion arouses great concern and it shows complex dynamical characteristics. The phenomenon of phase transitions has been found in the congested flow from the perspective of statistical mechanics and nonlinear dynamics. Moreover, the scientists and engineers are seeking solutions for improving the performance and safety of transportation system operation by the way of interdisciplinary studies. In this context, some great efforts have been made in the field of intelligent transportation systems (ITS), which has been regarded as the most effective method and advanced technology for reaching a more efficient and safe road usage.

The term ITS is used to describe a set of concepts and technologies for a kind of transportation systems while the main focus is on the integration of information and communication technologies (ICT) with infrastructure, vehicles, and traffic users. ITS aims to provide some innovative services for establishing harmonious and smart transportation system in which various traffic users can travel more safer and faster. A typical feature of ITS is the versatile nature of new technique. At present, wireless communications, computational technologies, sensing technologies, and pattern recognition have played a crucial role in the development of

ITS. Additionally, predictive techniques have been developed to provide advanced service for participants in transportation system. Of the various subsystems of ITS, for instance, advanced driver assistance system (ADAS) can improve efficiency and safety of vehicle operation by compensating for the unfavorable human behavior.

Nowadays, the concentration on low carbon, energy saving, and pollution control is increasing in the field of ITS. There are a number of beneficial forms of such green transportation that make our lives easier and more fun. Green transportation involves effective and efficient resource utilization, changes in transportation structure, and improving transportation environment. Green ITS will be the objective of development of transportation system in the future. Green ITS can make a positive contribution to the social, economic, and environmental sustainability in urban system. It can also provide harmonious connections between human and vehicles as well as infrastructures.

The focus of this special issue is to present the new achievements in basic and applied research relating to ITS and reveal the complex regularity of transportation system. The aim of this special issue is to contribute importantly to the development of green ITS, especially for the strategies for improving traffic performance and safety. This special issue invites original research from the theoretical and practical point of view. The topics of this special issue include the following:

intelligent transportation system,

discrete and dynamic problems in transportation system,

application of chaos in traffic flow,  
driving behavior and driver assistance system,  
discrete optimization methods in ITS,  
transit and rail systems operation,  
environment protection and energy saving in transportation,  
transportation policy and economics.

*Wuhong Wang*  
*Geert Wets*  
*Heiner Bubb*  
*Huimin Niu*

## Research Article

# Passenger Flow Prediction of Subway Transfer Stations Based on Nonparametric Regression Model

Yujuan Sun,<sup>1</sup> Guanghou Zhang,<sup>2</sup> and Huanhuan Yin<sup>3</sup>

<sup>1</sup> College of Mechanical Engineering and Applied Electronics Technology, Beijing University of Technology, Beijing 100124, China

<sup>2</sup> Institute of Comprehensive Transportation of NDRC, Beijing 100038, China

<sup>3</sup> Research Institute of Highway Ministry of Transport, Beijing 100088, China

Correspondence should be addressed to Yujuan Sun; yjsun@bjut.edu.cn

Received 10 November 2013; Revised 30 March 2014; Accepted 2 April 2014; Published 24 April 2014

Academic Editor: Huimin Niu

Copyright © 2014 Yujuan Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Passenger flow is increasing dramatically with accomplishment of subway network system in big cities of China. As convergence nodes of subway lines, transfer stations need to assume more passengers due to amount transfer demand among different lines. Then, transfer facilities have to face great pressure such as pedestrian congestion or other abnormal situations. In order to avoid pedestrian congestion or warn the management before it occurs, it is very necessary to predict the transfer passenger flow to forecast pedestrian congestions. Thus, based on nonparametric regression theory, a transfer passenger flow prediction model was proposed. In order to test and illustrate the prediction model, data of transfer passenger flow for one month in XIDAN transfer station were used to calibrate and validate the model. By comparing with Kalman filter model and support vector machine regression model, the results show that the nonparametric regression model has the advantages of high accuracy and strong transplant ability and could predict transfer passenger flow accurately for different intervals.

## 1. Introduction

Most cities in China are facing serious traffic problems, such as traffic congestion, pollution, and accidents. It is agreed that subway system is one of the efficient countermeasures to solve traffic problems. However, passenger flow is increasing dramatically with accomplishment of subway network system in big cities. As convergence nodes of subway lines, transfer stations need to assume more passengers due to amount transfer demand among different lines. Transfer facilities have to face great traffic pressure because passengers always arrive in a very short time. Consequently, pedestrian congestion or other abnormal situations will occur more easily. So, in order to avoid pedestrian congestion or warn the management before it occurs, it is very necessary to predict the transfer passenger flow to forecast pedestrian congestions.

Nonparametric regression was selected as the prediction method to forecast the passenger flow due to the fact that the authors have demonstrated the advantages of nonparametric regression over other approaches, such as Kalman filtering

[1, 2] and neural networks [3, 4] in previous research efforts, based on sufficient history data.

Nonparametric regression is suitable for uncertain and nonlinear dynamic system. It is founded on chaotic system theory. Earlier work by Smith [5] found that a simple implementation of the nearest neighbor forecasting approach provided reasonably accurate traffic condition forecasts. In 1987, Yakowitz [6] suggested the using of  $K$ -nearest neighbor method in time series forecasting. The basic approach of nonparametric regression is heavily influenced by its roots in pattern recognition [7]. In essence, the approach locates the state of the system (defined by the independent variables) in a neighborhood of past, similar states. Once this neighborhood has been established, the past cases in the neighborhood are used to estimate the value of the dependent variable.

Nonparametric regression model is quite suitable for deterministic and nonlinear prediction. And it could be used in the situation without transcendental knowledge and enough historical data. It can try to find the nearest neighbor between historical data and current data, and with the nearest neighbor, it tries to predict the flow in the next interval.

The algorithm assumes that the intrinsic links of all factors are all contained in the historical data. So, the information can be obtained directly from the historical data instead of establishing an approximate model for it. In other words, the nonparametric modeling does not smooth the historical data. Therefore, the predicted effect is more precise than the parameters modeling, especially in the special events. As a free parameter, portable, and high prediction accuracy algorithm, the error of nonparametric regression is relatively small. What is more, this model is quite suitable for computer programming and can be applied to the complex environment.

The basic idea of nonparametric regression is to form a typical historical database, which is on the basis of comprehensive analysis of a large number of historical data. The historical database contains variety of traffic state trends as well as the typical rules. Each type of data in the sample library represents a traffic evolution trend. The latest traffic data collected in real-time are matched with historical data to find the nearest  $K$ -group data. The prediction of coming traffic state is determined by the nearest neighbor trends of the  $K$ -group data. Accordingly, the whole algorithm has no fixed parameters and coefficients. It can predict the next period traffic state totally based on the sample database evolution trend and the value of real-time data. Historical data series are the typical mode of traffic evolution, which play an important role in the short-term prediction. Figure 1 shows the principle of nonparametric regression theory.

Due to well prediction ability, kinds of nonparametric regression models were used to forecast traffic states gradually. In 1991, Davis and Nihan [8] used the nonparametric regression in traffic forecasting. In 1997, Smith and Demetsky [9] used the last 5 months' data to forecast the traffic flow. The definition of state vector included historical average flows; the results were better than historical average and neural network methods. Oswald et al. [10] researched how to speed up the runtime of nonparametric regression, but the accuracy was degraded. Qi and Smith [11] developed a distance metric that can be effectively used with categorical data which commonly make up traffic event data. The metric was based on the influence of variable values on a measurable objective to the purpose of selecting the nearest neighbors. When this method was incorporated in a nonparametric regression forecasting model, it was demonstrated to outperform parametric forecasting models significantly.

Tang and Gao [12] enhanced the automatic incident detection ability for forecasting traffic flows based on improved nonparametric regression algorithms and standard deviation algorithms. Turochy [13] coupled nonparametric regression with a condition monitoring method which characterized the extent to which the current traffic conditions deviate from those that may be expected based on historical data. The mean absolute percentage errors for two of the four nearest neighbor forecasting procedures were reduced. Kindzerske and Ni [14] introduced a composite approach based on nonparametric regression which was used to predict traffic conditions. The composite approach performed the nearest neighbor search for each loop detector station only using the data which are in proximity to the detector's

position on the roadway. This method accommodated every detector station individually to minimize the forecast error on the entire roadway. And the composite approach can predict the onset and propagation of traffic shock waves.

Liu et al. [15] proposed a recursive nonparametric regression model and implemented it to forecast traffic flows and queue evolution in a congested actuated intersection. The model can be used to substitute traditional simulation software in the lower level of a real-time traffic control system to search the optimal control variables and then utilize the found solutions as the inputs in the simulation software in the upper level of that control system to attain the system performances. Shi and Ren [16] proposed a new method called MW model to improve the accuracy and computing speed of the nonparametric regression model when the database was too large and hard to search in short-term traffic flow forecasting. Zhang et al. [17] proposed a rule-based  $K$ -nearest neighbor nonparametric regression model to forecast large scale traffic flow of urban road networks. Rules were extracted from the historical data using Rough Set Theory, which assisted in finding the near neighbors.

Sun and Zhang [18] also proposed a selective random subspace predictor (SRSP) which was very similar to nonparametric regression model. The SRSP built selective input space based on Pearson correlation coefficients and then generated random input subspace to forecast. The method which the SRSP used to select relative variable could be used in nonparametric regression model.

From the previous literature review, it can be found that kinds of nonparametric regression models were widely used to predict traffic condition of motor vehicles. However, there were few research works related with pedestrian traffic. So, in order to test and verify the applicability of nonparametric regression in pedestrian traffic condition prediction, the  $K$ -nearest neighbor nonparametric regression model was used to forecast the transfer passenger flow of subway stations. The nonparametric regression's advantages of high accuracy and strong transplant ability are showed while being compared with Kalman filter model and support vector machine regression model.

## 2. Procedure of Nonparametric Regression Prediction

The application of nonparametric regression prediction contains five key steps: choosing clustering methods of historical database, the definition of state vector, the determining of similar mechanism, the choosing of the nearest neighbor mechanism, and the choosing of prediction function.

*2.1. Choosing Clustering Methods of Historical Database.* The first and critical step in nonparametric regression is historical data preparation, whose quality directly determines the prediction effect of nonparametric regression. What is more, the prediction effect of nonparametric regression is closely related to the choosing of clustering methods and computational time. Therefore, firstly, in order to search enough nearest neighbors, the historical database which was

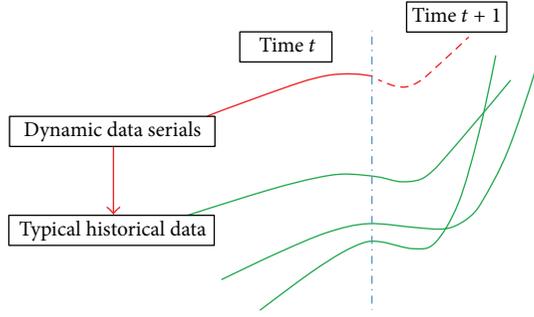


FIGURE 1: The schematic illustration of nonparametric regression theory.

built by clustering method must cover all state of the system. Secondly, clustering method should be able to meet the requirements in the dynamic data real-time classification and to meet the requirements of real-time, online programming. But now, traditional clustering methods take the average state vector or a single historical value as the clustering objects; it is difficult to reflect the data changing trends characteristics. Thus, the paper will focus on discussing the improvement of clustering methods and the model computational speed.

**2.2. Definition of State Vector.** State vector is composed of the minimum number of state variables, which are associated with the predictor variables. Because maybe there are a lot of state variables associated with predictor variables, it is necessary to properly select the number of state vectors to achieve the best balance between accuracy and computational speed.

**2.3. Similar Mechanism.** It is an important concept in the nonparametric regression, which means how to evaluate the similarity of the current point and the historical database. The most commonly used metric method is the Euclidean distance or weighted Euclidean distance.

**2.4. Choosing the Nearest Neighbor Mechanism.** As a core concept of nonparametric regression, the nearest neighbor mechanism refers to the point in the history database and how to become a close neighbor of the current point. There are two mechanisms: minimum  $K$ -nearest neighbor method and nuclear nearest neighbor method, respectively. The minimum  $K$ -nearest neighbor method means  $K$  points, whose similarity is the biggest in historical database. The nuclear nearest neighbor method refers to taking the current point as the core; all points within the radius of  $R$  become the nearest neighbor of the current point.

**2.5. Prediction Function Selection.** After finding the nearest neighbor points, a function needs to be used to take advantage of these points to predict the next period value. Commonly used methods are average, weighted average, and so on.

### 3. Improvement of Typical Model

**3.1. Improvement of Historical Data Clustering.** The basic procedure of nonparametric regression prediction is to compare the recent data status with the historical data and figure out the most similar data serials which would be used to predict the future data status. So, in order to provide the most similar data serial, the historical database should include enough historical information. And, in order to reflect as many trends of data serial as possible, all the historical data were stored in the database without any processing. So, the organization method of data serial in historical database determines the calculation efficiency of the prediction model. The historical database is the foundation of transfer passenger flow prediction. The core concept of the nonparametric regression is to match recent data with the historical database. From all the matches, either the  $K$  nearest matches or all the matches below a given distance threshold are located. According to the data storing system of computer science, an improved historical data organization method is proposed. This method quantifies the trend of historical data serial and sets different value for different trend which is used to cluster the historical data serials.

If the length of the data serials is  $n$ , then the historical data serial is  $S_h(t) = \{S_h(t-n+1), S_h(t-n+2), \dots, S_h(t)\}$ , and the recent data serial is  $S(t) = \{S(t-n+1), S(t-n+2), \dots, S(t)\}$ . Thus, the next data serial of historical database and recent status are  $S_h(t+1)$  and  $S(t+1)$ , respectively.

If  $d = \{0, 1, 2\}$  is the trend description serial of data serial, then the value of the trend description serial is

$$D_{\text{label}}(i) = \begin{cases} 0 & s(t-n+i) = s(t-n+i+1) \\ 1 & s(t-n+i) < s(t-n+i+1) \\ 2 & s(t-n+i) > s(t-n+i+1), \end{cases} \quad (1)$$

$$i = 1, 2, \dots, n-1.$$

The number of clustering types of historical database is

$$C_{\text{no}} = 3^{n-1}. \quad (2)$$

For one data serial, the clustering label is

$$C_{\text{label}} = D_{\text{label}}(1) \times 3^{n-2} + \dots + D_{\text{label}}(n-2) \times 3^1 + D_{\text{label}}(n-1) \times 3^0. \quad (3)$$

Figure 2 is the trend of one data serial with length of 4. Based on (2), the number of clustering types in historical database is

$$C_{\text{no}} = 3^{4-1} = 27. \quad (4)$$

And the clustering label is

$$C_{\text{label}} = 1 \times 3^2 + 1 \times 3^1 + 2 \times 3^0 = 14. \quad (5)$$

**3.2. The Selection of Data Serial.** Based on the experimental analysis, the neighbor data are chosen as the state vector. The vector contains four current transfer passenger flow trend

TABLE 1: Precision of nonparametric regression forecasting model.

Performance	Time					
	7:00–9:00			17:00–19:00		
	1 minute	3 minutes	5 minutes	1 minute	3 minutes	5 minutes
Average relative error	12.20%	8.10%	6.30%	11.80%	6.00%	4.00%
Maximum relative error	42.00%	35.00%	23.00%	31.00%	24.00%	13.00%
Equalization coefficient	0.91	0.96	0.96	0.93	0.96	0.98

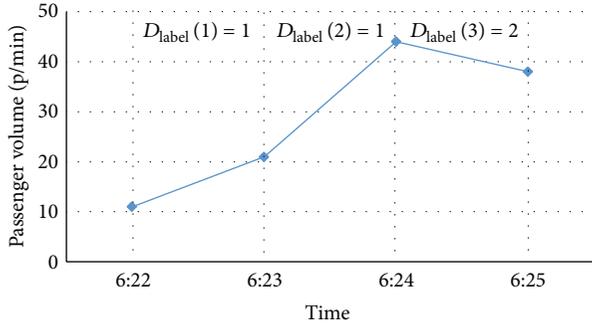


FIGURE 2: Illustration of trend label of state vector for nonparametric regression.

data and five historical transfer passenger flow trend data. Four neighbor data are selected as data serial. The prediction model calculates the clustering label based on the trend of the four neighbor data and searches for the most similar data serials from history database. Then, the future data status is predicted according to the next trend of the most similar data serials.

**3.3. The Similar Mechanism.** The Euclidean distance is used to calculate the similar level between the recent data serial and the historical data serials. The equation is

$$d_i = \left( \frac{1}{4}(S(t-3) - S_h(t-3))^2 + \frac{1}{4}(S(t-2) - S_h(t-2))^2 + \frac{1}{4}(S(t-1) - S_h(t-1))^2 + \frac{1}{4}(S(t) - S_h(t))^2 \right)^{1/2}. \quad (6)$$

Except for the Euclidean distance, the weights of the most similar historical data serials are also used in the prediction model. As shown in (7),  $\beta_i$  is the weight of the most similar historical data serial  $i$ . The bigger the  $\beta_i$  is, the more remarkable the influence level on the prediction result of data serial  $i$  is:

$$\beta_i = \frac{d_i}{\sum_{j=1}^k d_j}, \quad (7)$$

where  $k$  is the number of the most similar data serials.

**3.4. The Selection of Neighbor Mechanism.**  $K$ -neighbor mechanism is used to select the nearest neighbors.  $K$  represents for

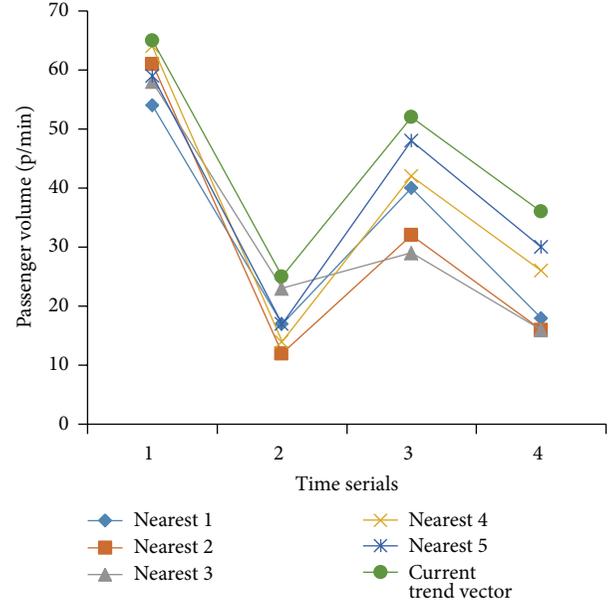


FIGURE 3: Comparison of state vector of prediction and similar neighborhood.

the numbers of nearest neighbors which are selected from historical database, and has close relation to the database's character. Based on the previous research results [13, 14, 19], the  $K$  is 5.

**3.5. The Improvement of Selection Model.** The weighted average method based on the reciprocal of the matching distance is chosen as the prediction function. The shorter distance point is the more similar point. Then, the weighing is bigger. For most nonparametric regression prediction models, the next value of the most similar historical serial is used as the prediction value of recent data serial. The next value and weighted coefficient based on the historical data are used to predict the transfer passenger flow in the prediction algorithm. In the state vector of the prediction model, the historical data of the current time and the nearest time are used to identify different prediction coefficient, and the historical data of the next trend are used to calculate the prediction data directly.

However, due to reasons such as the lack of historical data or abnormal flow, the next value of recent data serial may change dramatically, taking Figure 3 as an example. So, in order to improve the prediction accuracy, the amending

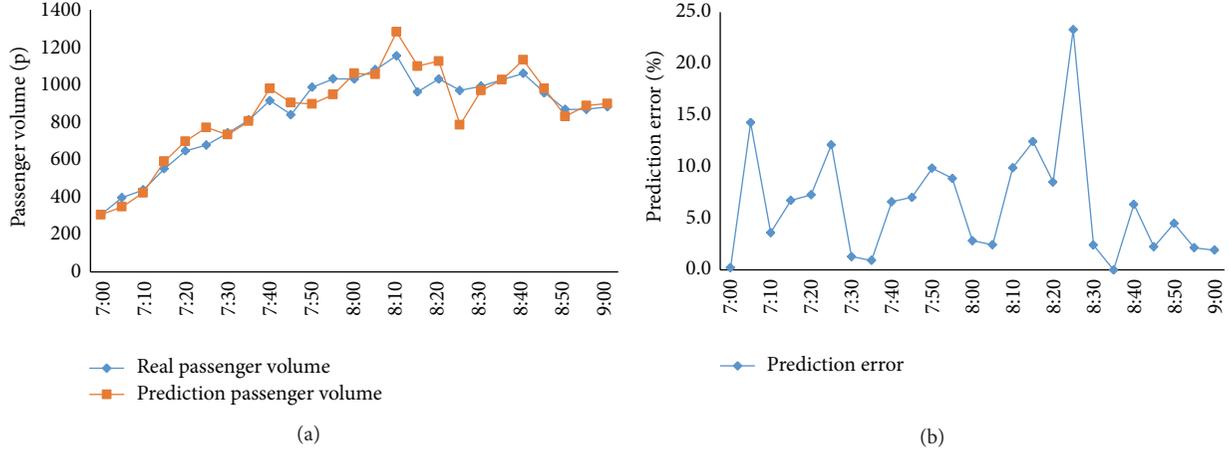


FIGURE 4: Forecasting result for each 5 minutes in morning peak hour using nonparametric regression.

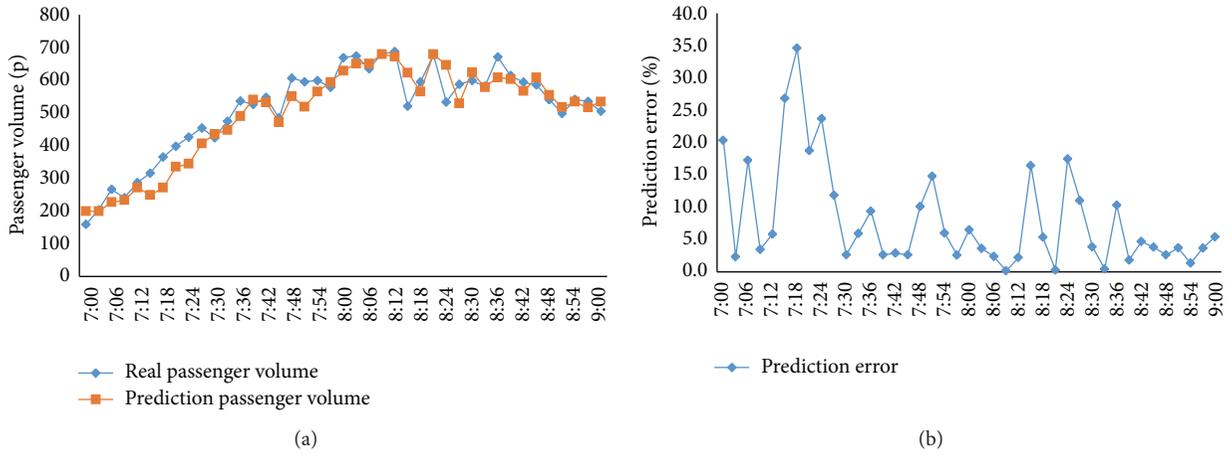


FIGURE 5: Forecasting result for each 3 minutes in morning peak hour using nonparametric regression.

coefficient with average value of recent data serial is proposed. The improved model is

$$s(t+1) = \frac{\sum_{i=1}^K \beta_i s_{hi}(t+1) \bar{x}}{\bar{x}_{hi}}, \quad (8)$$

where  $\bar{x}$  is the average value of recent data serial given as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n s(t-i+1) \quad (9)$$

and  $\bar{x}_{hi}$  is the average value of the neighbor data serial  $i$  given as

$$\bar{x}_{hi} = \frac{1}{n} \sum_{j=1}^n s_{hi}(t-j+1). \quad (10)$$

#### 4. Application

In order to test the accuracy of the prediction model, the transfer passenger flow of XIDAN station was used to

calibrate the model. The historical database was built with the transfer passenger flow from July 26 to August 25, 2011. The prediction data were the passenger flow of August 25, 2011. The prediction results are illustrated in Figure 4 to Figure 9.

4.1. Forecasting Results of Peak Hours from 7:00 to 9:00. See Figures 4, 5, and 6.

4.2. Forecasting Results of Peak Hours from 17:00 to 19:00. See Figures 7, 8, and 9.

The prediction performance for different time and intervals is shown in Table 1. It is obvious that the improved nonparametric regression model has very high prediction accuracy. The maximum average relative error is less than 15%. To compare the applicability of different prediction model, the Kalman filter model and support vector machine regression model are chosen to predict the transfer passenger flow. Figures 10, 11, and 12 show the comparison of prediction capability of three different models. Compared with the Kalman filter model and support vector machine regression model, the accuracy of the predicted transfer passenger flow

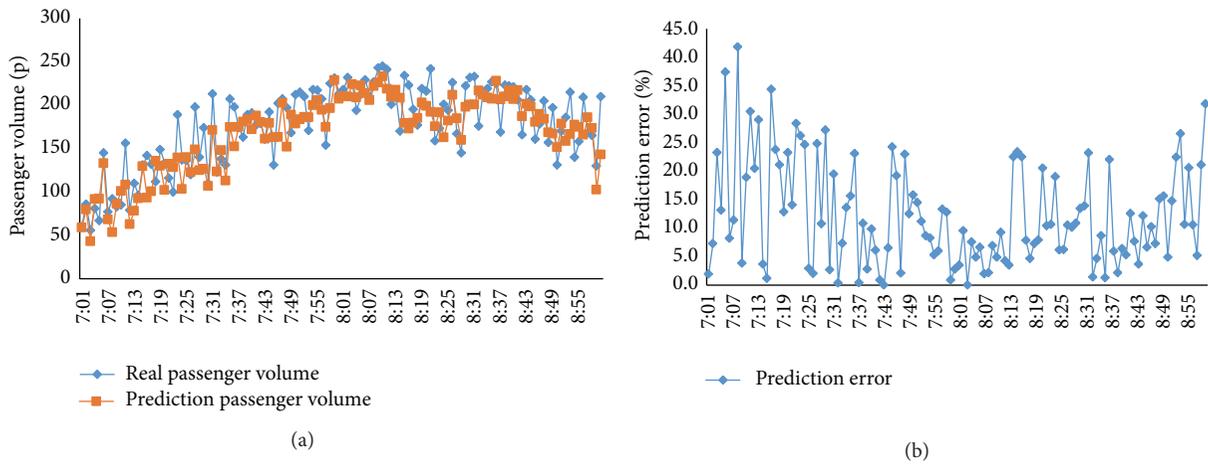


FIGURE 6: Forecasting result for each 1 minute in morning peak hour using nonparametric regression.

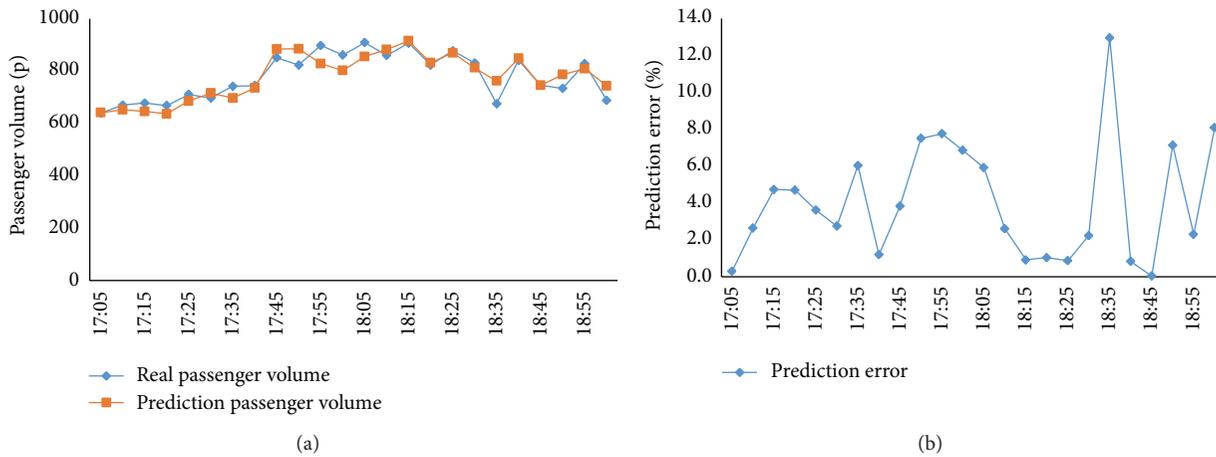


FIGURE 7: Forecasting the result for each 5 minutes in evening peak hour using nonparametric regression.

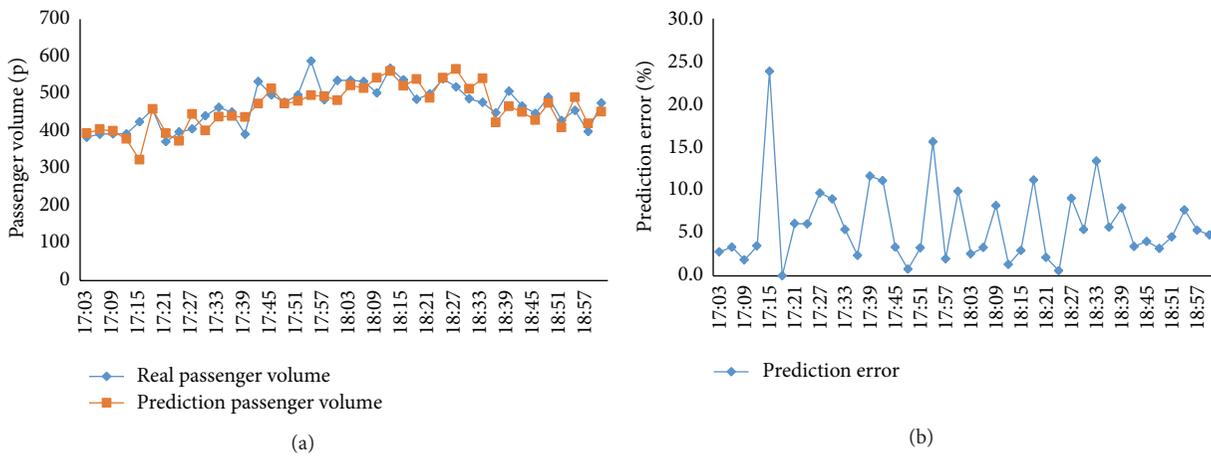


FIGURE 8: Forecasting the result for each 3 minutes in evening peak hour using nonparametric regression.

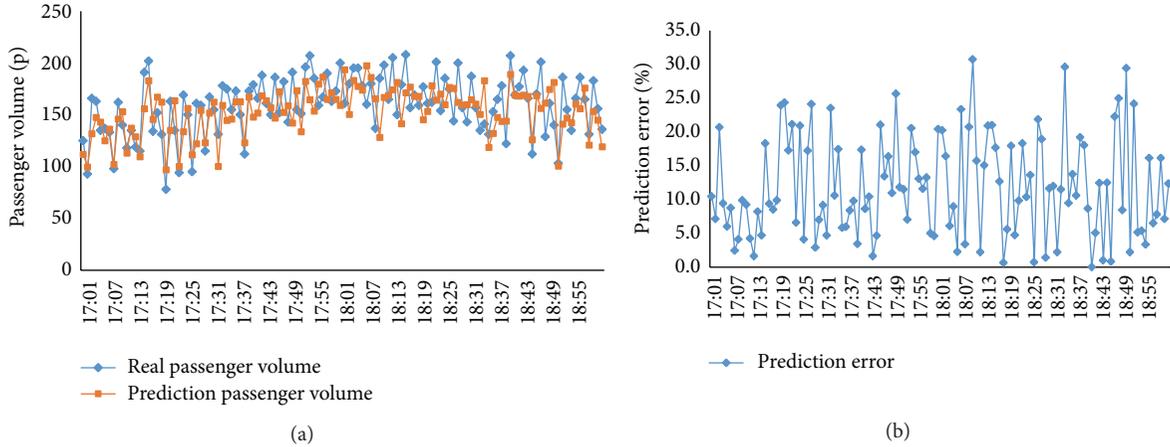


FIGURE 9: Forecasting the result for each 1 minute in evening peak hour using nonparametric regression.

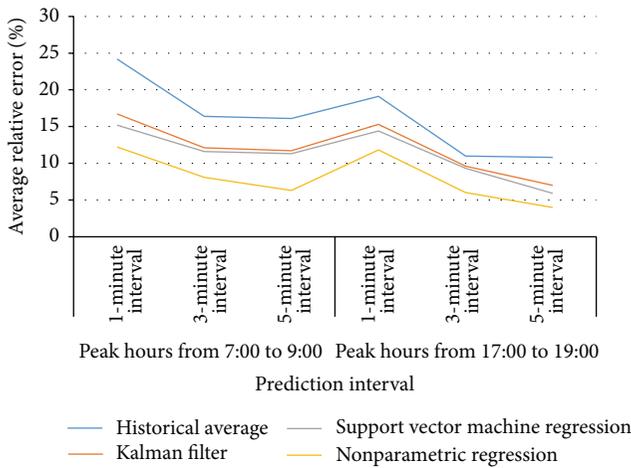


FIGURE 10: Comparison of average relative error for different forecasting models.

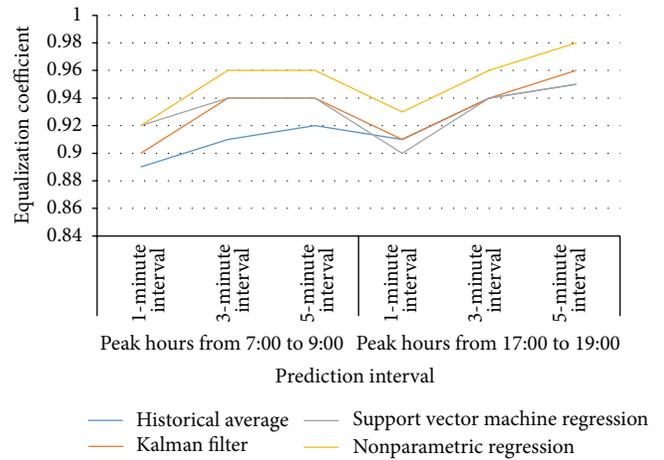


FIGURE 12: Comparison of equalization coefficient for different forecasting models.

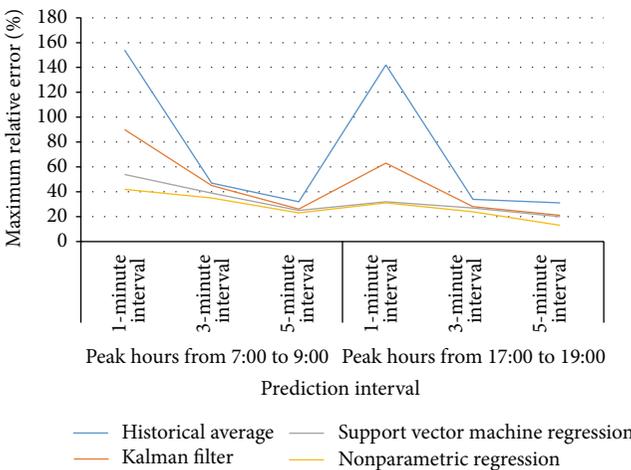


FIGURE 11: Comparison of maximum relative error for different forecasting models.

and the stability of the error for the improved nonparametric regression model have been improved significantly. So, the improved nonparametric regression prediction model can be used in real application.

### 5. Conclusions

As a convergence node of subway lines, transfer stations need to assume more passengers due to amount transfer demand among different lines. So, it is really very necessary to predict the transfer passenger flow to avoid pedestrian congestion or warn the management before it occurs.

Based on nonparametric regression theory, a transfer passenger flow prediction model was proposed. And data of transfer passenger flow for one month in XIDAN transfer station were used to calibrate and validate the model. The results show that the model could predict transfer passenger flow accurately for different intervals. What is more, the prediction accuracy is also much better than Kalman filter model and support vector machine regression model. The

bigger the interval is, the more accurate the prediction result is. The maximum average relative error is 12.20%, which means that the prediction model can be used in real application.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### References

- [1] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B*, vol. 18, no. 1, pp. 1–11, 1984.
- [2] J. Whittaker, S. Garside, and K. Lindveld, "Tracking and predicting a network traffic process," *International Journal of Forecasting*, vol. 13, no. 1, pp. 51–61, 1997.
- [3] S. C. Chang, S. J. Kim, and B. H. Ahn, "Traffic-flow forecasting using time series analysis and artificial neural network: the application of judgmental adjustment," in *Proceeding of the 3rd IEEE International Conference on Intelligent Transportation Systems*, Dearborn, Mich, USA, October 2000.
- [4] L. R. Rilett and D. Park, "Direct forecasting of freeway corridor travel times using spectral basis neural networks," *Transportation Research Record*, no. 1752, pp. 140–147, 2001.
- [5] B. L. Smith, *Forecasting freeway traffic flow for intelligent transportation system applications [Doctoral dissertation]*, Department of Civil Engineering, University of Virginia, Charlottesville, Va, USA, 1995.
- [6] S. Yakowitz, "Nearest-neighbour methods for time series analysis," *Journal of Time Series Analysis*, vol. 8, no. 2, pp. 235–247, 1987.
- [7] M. Karlsson and S. Yakowitz, "Rainfall-runoff forecasting methods, old and new," *Stochastic Hydrology and Hydraulics*, vol. 1, no. 4, pp. 303–318, 1987.
- [8] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting," *Journal of Transportation Engineering*, vol. 117, no. 2, pp. 178–188, 1991.
- [9] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: comparison of modeling approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266, 1997.
- [10] R. K. Oswald, W. T. Scherer, and B. L. Smith, "Traffic flow forecasting using approximate nearest neighbor nonparametric regression," Research Report Uvacts-15-13-7, Center for transportation studies at the University of Virginia, 2001.
- [11] Y. Qi and B. L. Smith, "Identifying nearest neighbors in a large-scale incident data archive," *Transportation Research Record*, no. 1879, pp. 89–98, 2004.
- [12] S. Tang and H. Gao, "Traffic-incident detection-algorithm based on nonparametric regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 38–42, 2005.
- [13] R. E. Turochy, "Enhancing short-term traffic forecasting with traffic condition information," *Journal of Transportation Engineering*, vol. 132, no. 6, pp. 469–474, 2006.
- [14] M. D. Kindzerske and D. Ni, "Composite nearest neighbor nonparametric regression to improve traffic prediction," *Transportation Research Record*, no. 1993, pp. 30–35, 2007.
- [15] K. Liu, R. Ghaman, and F. Xiang, "An adaptive procedure for prediction of traffic conditions at signalized intersection," in *Proceedings of the 14th World Congress on Intelligent Transport Systems (ITS '07)*, Beijing, China, October 2007.
- [16] X. Shi and Q. Ren, "The advanced nonparametric model for short-Term traffic volume forecasting," in *Proceedings of the 11th International Conference of Chinese Transportation Professionals (ICCTP '11)*, pp. 1442–1453, Nanjing, China, August 2011.
- [17] L. Zhang, Z. Peng, D. Sun, and X. Liu, "A novel rule-based traffic state forecasting approach for large-scale road networks," *Transportation Research Record*, vol. 2279, pp. 3–11, 2012.
- [18] S. Sun and C. Zhang, "The selective random subspace predictor for traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 367–373, 2007.
- [19] S. Clark, "Traffic prediction using multivariate nonparametric regression," *Journal of Transportation Engineering*, vol. 129, no. 2, pp. 161–168, 2003.

## Research Article

# Activity-Trip Chaining Behavior of Urban Low-Income Populations in Nanjing, China: A Structural Equations Analysis

Zhaoming Chu,<sup>1</sup> Hui Chen,<sup>2</sup> Lin Cheng,<sup>1</sup> Xuewu Chen,<sup>1</sup> and Senlai Zhu<sup>1</sup>

<sup>1</sup> School of Transportation, Southeast University, Nanjing 210096, China

<sup>2</sup> Chengxian College, Southeast University, Nanjing 210088, China

Correspondence should be addressed to Zhaoming Chu; chuzhaoming@126.com

Received 31 October 2013; Revised 22 March 2014; Accepted 27 March 2014; Published 22 April 2014

Academic Editor: Huimin Niu

Copyright © 2014 Zhaoming Chu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper analyzes the activity-trip chaining behavior of urban low-income populations in Nanjing, China, based on a specific travel survey of low-income residents of Nanjing city (2010), and the database of residents travel survey of Nanjing city (2009). Individual's information of activity participation and trip chains is extracted from the daily travel diary and matched with individual and household characteristics. On top of correlation analysis and normalization process, using the software AMOS, two structural equation models are formulated to analyze the relationship among individuals' sociodemographics, activity duration, and trip chains of low-income populations and non-low-income populations, respectively. Seven household characteristics and six individual characteristics are chosen as the exogenous variables, while 4 indices of activity duration and 4 indices of trip chains are selected as the endogenous variables. The result shows that the activity-travel behavior of urban low-income populations is quite unique, which offers promising insights into activity-trip chaining behavior of the poor and extends the need to crafting effective transportation policies specifically for urban low-income populations in developing countries.

## 1. Introduction

Because of the fast urbanization in developing countries, large amount of peasants swarm to the urban area and work there. These peasant-workers together with unemployed city residents consist the main parts of urban low-income populations in developing countries. However, in recent years, the soaring house price in big cities forces urban low-income populations move to the urban fringe in developing countries; meanwhile, transportation becomes a big problem to these low-income residents. How to satisfy the travel demand of urban low-income residents and how to provide cheap and convenient service for them are urgent problems to be solved. The research on travel behavior of low-income residents can help to promote social fairness and justice, ease the social conflicts, and build a harmonious society.

To solve the various transportation problems encountered by the urban low-income populations in developing countries, it is required to capture the characteristics of their activity-travel behavior first. An activity-based survey method was selected because it typically yields higher rates of

trip recall than other methods and is therefore relatively well suited to investigate travel behavior in its fuller complexity. As an extremely flexible linear-in-parameters multivariate statistical modeling technique, structural equation modeling (SEM) has been proved to have considerable potential in modeling activity-based travel demand modeling.

The primary intention of this paper is to explore the characteristics of activity-trip chaining behavior of urban-low-income populations in developing countries. Specifically, we use the structural equations to analyze the correlation of sociodemographic, activity participation, and trip chaining behavior of urban low-income populations and capture the most critical factors that influence the travel behavior of the poor by comparing the urban low income populations and non-low-income populations. Our findings are expected to further add to the rich body of knowledge on activity-based travel demand modeling by focusing on urban low-income populations, meanwhile providing useful information for crafting effective policies to guarantee the cheap and convenient travel of the poor.

The organization of the paper is as follows. In the following section, we briefly review the relevant literature on the topic of this study. In Section 3, we introduce the principle of structural equation modeling. Section 4 describes the data used in this paper. Section 5 presents the development and calibration of the models. Section 6 presents selected findings of model estimation results. Finally, the paper ends with conclusions and future research directions.

## 2. Literature Review

In the literature, travel behavior of urban low-income residents has not been studied much due to the limited data. Several representative studies are listed as follows.

Giuliano et al. examined the use of public transit by low-income households and they declared that public transit is not a reasonable substitute for the private vehicle for most people, poor or not poor [1, 2]. Blumenberg and Haas claimed that welfare recipients with unlimited access to automobiles have higher employment rates and report fewer transportation problems [3]. Clifton presented a few of the challenges facing those interested in the intersection between poverty and travel behavior and introduced opportunities to explore low-income travel using some new approaches [4]. McDonald et al. found that the free-bus pass program increased low-income students' bus ridership and after-school participation; they also found that the increases in bus use were greater among free-bus pass holders, in areas with high levels of bus service and among high school students [5].

The above studies are all based on the data of developed countries. Recently, some researches started to focus on the low-income populations in developing countries. Using an activity diary survey administered in Cape Town, a city of South Africa, Behrens found that travel occurring by nonmotorized modes, for non-work purposes and during off-peak periods, is considerable. They also argued that restricting the focus of analysis to motorized, work, and peak period trip-making can create serious misconceptions of the true nature of travel behavior, particularly of low-income households [6]. Srinivasan and Rogers surveyed 70 households which were located in two different parts of Chennai (a city of India). The results indicated that residents in the centrally located settlement were more likely to use nonmotorized travel modes than the peripherally located residents [7].

More recently, researchers paid more attention to the mobility of low-income populations. Thakuriah et al. proposed an index of perceived service importance (PSI) to evaluate low-income transit services [8]. Taylor et al. focused on the role of the car of the travel choices and needs of low-income households and they concluded that the car clearly plays an important role in the lives of low-income households [9]. Gao and Johnston examined possible impacts of car ownership promotion versus transit improvements on job accessibility, work trips, and traveler benefits, for low-income households [10].

However, most of the travel behavior studies pertaining to low-income populations are trip-based and ignored the effects of activity participation on travel behavior. None of

these studies have examined trip chaining behavior of the poor.

After 30 years of development, activity-based travel demand modeling has been widely used in travel behavior analysis, but there are few activity-based travel demand models specifically for low-income populations. Meanwhile, due to the availability of improved software, structural equation modeling (SEM) has become an effective tool to model travel behavior, especially in the field of activity-based travel demand modeling [11].

Kitamura et al. was the first to apply SEM in modeling joint demand for activity duration and travel. Based on the California time use survey data, they confirmed a negative feedback of commute time to non-work activities [12]. Lu and Pas described the development, estimation, and interpretation of a model relating sociodemographics, activity participation, and travel behavior. Using the structural equation modeling methodology with the endogenous variables of travel behavior indices, a complex set of interrelationships among the variables of interest is estimated simultaneously. They found that travel behavior can be better explained by including activity participation endogenously in the model than through sociodemographics alone [13]. Golob estimated a joint model of work and non-work activity duration, four types of trip chains, and three measures of travel time expenditure. In this model, maximum likelihood (ML) estimation was applied to Portland data, and the effects of in-home work and residential accessibility were also explored [14].

More recently, using activity-based travel survey data collected in the Washington, DC metropolitan area, Kupam and Pendyala carried out an exploratory analysis of commuters' activity and travel patterns to investigate and estimate relationships among sociodemographics, activity participation, and travel behavior. The model estimation results show that significant trade-offs exist between in-home and out-of-home activity participation [15]. Chung and Ahn used structural equation models to analyze the day-to-day activity participation and travel behavior in a developing country. They confirmed that activity patterns are significantly different on weekdays and weekends. Furthermore, they found that during weekdays there are some day-to-day variations in the patterns of activity participation and travel behavior [16].

Due to the modern and hectic life style, travel behavior of people is becoming complex day by day, especially in fast developing countries. Therefore, the better understanding of trip chain decision making is necessary to transportation researchers and policy makers. In the last few years, researchers began to study the trip-chaining behavior in developing countries [17, 18]. However, to the best of our knowledge, there is still no research specifically for the low-income populations in developing countries, who are the main focuses of this study.

To sum up, previous studies have confirmed that activity participation had a significant relationship with travel behavior, particularly in time use. However, less work has been done to explore the impact of activity participation on trip chaining specifically for low-income populations. Since previous studies have primarily focused on households in developed

courtiers, there is a great need to formulate relationships of sociodemographics, activity participation, and trip chaining of urban low-income populations in developing countries, such as China.

### 3. Methodology

In order to estimate a simultaneous model of the inter-relationship among sociodemographics, activity duration, and trip chaining behavior of urban low-income residents, we applied the methodology of structural equation model (SEM). In addition, we are also interested in the direct and indirect effects of one variable on another, which can be provided by the estimation result of structure equation model.

Since all variables used in this research are observed variables, structural equation models without latent variables are therefore reduced to the following form:

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \mathbf{\Gamma}\mathbf{x} + \boldsymbol{\zeta}, \quad (1)$$

where  $\mathbf{y}$  is a column vector of  $p$  endogenous variables,  $\mathbf{x}$  is a column vector of  $q$  exogenous variables,  $\mathbf{B}$  is a matrix ( $p \times p$ ) of direct effects between pairs of  $p$  endogenous variables,  $\mathbf{\Gamma}$  is a matrix ( $p \times q$ ) of regression effects associated with exogenous variables, and  $\boldsymbol{\zeta}$  is a column vector of the error terms, with the standard assumption that  $\boldsymbol{\zeta}$  is uncorrelated with  $\mathbf{x}$ . Further, we denote  $\boldsymbol{\Phi}$  by the covariance matrix of  $\mathbf{x}$  and  $\boldsymbol{\Psi}$  by the covariance matrix of  $\boldsymbol{\zeta}$ .

Structural equations systems are estimated by covariance-based structural analysis, in which the difference between the sample covariance and the model implied covariance matrices is minimized. The fundamental hypothesis for the covariance-based estimation procedures is that the covariance matrix of the observed variable is a function of a set of parameters as shown in the following equation:  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ , where  $\boldsymbol{\Sigma}$  is the population covariance matrix of observed variables,  $\boldsymbol{\theta}$  is a vector that contains the model parameters, and  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  is the covariance matrix written as a function of  $\boldsymbol{\theta}$ .

The matrix  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  has three components, namely, the covariance matrix of  $\mathbf{y}$ , the covariance matrix of  $\mathbf{x}$  with  $\mathbf{y}$ , and the covariance matrix of  $\mathbf{x}$ . Then, it can be shown that

$$\begin{aligned} \boldsymbol{\Sigma}(\boldsymbol{\theta}) &= \begin{bmatrix} \boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{yx}(\boldsymbol{\theta}) \\ \boldsymbol{\Sigma}_{xy}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}) \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{I} - \mathbf{B})^{-1} (\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})(\mathbf{I} - \mathbf{B})^{-1'} & (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}'(\mathbf{I} - \mathbf{B})^{-1'} & \boldsymbol{\Phi} \end{bmatrix}. \end{aligned} \quad (2)$$

The unknown parameters in  $\mathbf{B}$ ,  $\boldsymbol{\Gamma}$ ,  $\boldsymbol{\Phi}$ , and  $\boldsymbol{\Psi}$  are estimated so that the implied covariance matrix  $\boldsymbol{\Sigma}$  is as close as possible to the sample covariance matrix  $\mathbf{S}$ . In order to achieve this, a fitting function  $\mathbf{F}(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ , which is to be minimized, is defined. The fitting function has the properties of being scalar, greater than or equal to zero if and only if  $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{S}$ , and continuous in  $\mathbf{S}$  and  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  [19].

Several methods can be used to estimate the parameter in structural equation model, including: maximum likelihood (ML), unweighted least squares (ULS), generalized

least squares (GLS), and diagonally weighted least squares (DWLS). In this paper, we primarily used the ML estimation approach.

### 4. Data Description

The city selected for this study is Nanjing, the capital of Jiangsu Province, China. With a total land area of 6,589 square kilometers and an urban population of over eight million (2013), Nanjing is the second largest commercial center in East China after Shanghai.

There are two sources of data; the group of low-income population is from a specific travel survey of low-income residents of Nanjing City (2010) and the group of non-low-income is from the database of residents travel survey of Nanjing City (2009). In both surveys, all respondents were asked to record their activity and travel information within one weekday on a travel diary. In addition to the activity and travel information, each respondent is required to report the usual set of his/her household and personal sociodemographics.

In the specific survey of low-income populations (2010), 1000 questionnaires were delivered to low-income people residing in three parts of Nanjing City, including shanty areas in inner-city (300 copies), welfare-oriented public housing neighborhoods in the edge area of inner-city (200 copies), and the economically affordable housing neighborhoods in urban fringe (500 copies). Then, totally 904 questionnaires returned from all the surveyed areas.

The non-low-income group consists of residents whose annual per-capita income is higher than the minimum salary threshold of Nanjing City. Thus, 8666 non-low-income residents are selected from the database of residents travel survey of Nanjing City (2009). After eliminating missing data and performing logic checking, we selected 846 individuals in the low-income group and 7534 individuals in the non-low-income group.

Based on previous research and single factor analysis of sociodemographic attributes and endogenous variables, 7 household attributes and 6 individual attributes are selected as the exogenous variables, while the indices of activity participation and trip chains are selected as the endogenous variables. In particular, the descriptors of activity participation are defined by the duration of four types of activity: in-home, subsistence, maintenance, and leisure. The trip chaining characteristics are defined by descriptors of 4 items, namely, number of work chains, travel time of work chains, number of non-work chains, and travel time of non-work chains (see Table 1).

Statistical characteristics of exogenous variables are shown in Tables 2 and 3. It can be found that 67.5% of low-income residents live in the urban fringe, while 60.1% of the non-low-income residents live in main urban area. Annual household income of low-income populations mainly concentrates on low groups of 10,000~20,000, and 20,000~50,000 RMB, which take up 32.6% and 45.0%, respectively. In contrast, their non-low-income counterparts concentrate on middle-to-high groups of 20,000~50,000 and 50,000~

TABLE 1: Endogenous variables and exogenous variables.

	Variable	Label	Notes	
Exogenous variables	Residential location	Big_zone	Main urban area = 1, urban fringe = 2	
	Number of family members	N_people		
	Number of preschool children	N_kid		
	Household characteristics	Annual household income	Income	
		Number of vehicles	N_car	
		Number of bikes	N_bike	
		Number of electric bicycles	N_ebike	
		Gender	Sex	Male = 1, female = 2
		Job	Job	9 categories
		Transit IC card holding	IC	Hold a bus IC card = 1, other = 0
		Age	Age	8 categories
		Driving license holding	Lic	Hold a driver's license = 1, other = 0
		Educational level	Edu	4 categories
Endogenous variables	Activity duration	In-home activity	D_H	Sleeping, dinner, housework, and so forth
		Subsistence	D_S	Work, work-related and school
		Maintenance	D_M	Obligations, and so forth
		Leisure	D_L	Amusement, exercise, relaxation, and so forth
	Trip chaining	Number of work chains	N_W	Number of work related chains per day
		Travel time of work chains	T_W	Travel time of work chains per day
		Number of non-work chains	N_O	Number of non-work chains per day
		Travel time of non-work chains	T_O	Travel time of non-work chains per day

TABLE 2: Statistical characteristics of household characteristics.

Variable		Low-income group			Non-low-income group		
		Cases	Valid percent	Cumulative percent	Cases	Valid percent	Cumulative percent
Big_Zone	Inner city	275	32.5	32.5	4526	60.1	60.1
	Urban fringe	571	67.5	100.0	3008	39.9	100.0
N_people	1	15	1.8	1.8	39	0.5	0.5
	2	104	12.3	14.1	1489	19.8	20.3
	3	444	52.5	66.5	5076	67.4	87.7
	≥4	283	33.4	100.0	921	12.3	100.0
N_Kid	0	632	74.7	74.7	6613	87.8	87.8
	1	200	23.6	98.3	892	11.8	99.6
	≥2	14	1.7	100.0	29	0.4	100.0
Income	<¥10,000	38	4.5	4.5	0	0	0
	¥10,000~¥20,000	276	32.6	37.1	0	0	0
	¥20,000~¥50,000	381	45	82.2	3962	52.6	52.6
	¥50,000~¥100,000	151	17.8	100.0	2605	34.6	87.2
	>¥100,000	0	0	100.0	967	12.8	100.0
N_Car	0	692	81.8	81.8	5790	76.9	76.9
	1	150	17.7	99.5	1640	21.8	98.6
	≥2	4	0.5	100.0	104	1.4	100.0
N_bike	0	225	26.6	26.6	1654	22.0	22
	1	507	59.9	86.5	3315	44.0	66.0
	2	104	12.3	98.8	1988	26.4	92.3
	≥3	10	1.2	100.0	577	7.7	100.0
N_ebike	0	343	40.5	40.5	3330	44.2	44.2
	1	347	41	81.6	3314	44	88.2
	2	140	16.5	98.1	833	11.1	99.2
	≥3	16	1.9	100.0	57	0.8	100.0

TABLE 3: Statistical characteristics of individual characteristics.

Variable	Low-income group			Non-low-income group			
	Cases	Valid percent	Cumulative percent	cases	Valid percent	Cumulative percent	
Gender	Male	416	49.2	49.2	1	49.7	49.7
	Female	430	50.8	100	2	50.3	100.0
Job	School children	73	8.6	8.6	1	11.3	11.3
	College student	15	1.8	10.4	2	2.6	13.9
	Factory worker	124	14.7	25.1	3	15.0	28.8
	Service staff	114	13.5	38.5	4	8.9	37.7
	Civil servant	79	9.3	47.9	5	25.8	63.5
	Self-employed	43	5.1	53.0	6	5.8	69.3
	Retired	246	29.1	82.0	7	17.7	87.0
	Peasant	51	6.0	88.1	8	1.3	88.3
	Others	101	11.9	100.0	9	11.7	100.0
Transit IC Card	Yes	722	85.3	85.3	1	63.9	63.9
	No	124	14.7	14.7	2	36.1	100.0
Age	6~14	43	5.1	5.1	1	6.6	6.6
	15~19	30	3.5	8.6	2	5.1	11.7
	20~24	75	8.9	17.5	3	5.6	17.2
	25~29	115	13.6	31.1	4	8.8	26.1
	30~39	127	15.0	46.1	5	19.4	45.5
	40~49	110	13.0	59.1	6	24.3	69.8
	50~59	148	17.5	76.6	7	18.6	88.4
	≥60	198	23.4	100	8	11.6	100.0
Driving license	Yes	118	13.9	13.9	1	27.6	27.6
	No	728	86.1	100.0	2	72.4	100.0
Educational level	Middle school	429	50.7	50.7	1	26.5	26.5
	High School	323	38.2	88.9	2	38.5	64.8
	Undergraduate	94	11.1	100.0	3	33.9	98.7
	Graduate				4	1.3	100.0

100,000 RMB. Note that, 18.2% of low-income household own at least one car and 13.9% of the low-income individuals hold a driving license, which indicates that automobile begin to enter the Chinese urban families, even the not so affluent ones.

Table 4 shows statistical characteristics of the 8 endogenous variables that consist of descriptors of activity and trip chaining. Note that, on average, the duration of out-of-home activities are less in the low-income group than that of the non-low-income group. The number of trip chains indicates that low-income populations generally do less out-of-home activities.

### 5. Model Specification

On the basis of activity-based travel demand theory and previous researches on SEMs, a possible structural equation modeling framework was laid out as shown in Figure 1, which captures the interrelationships among sociodemographics, activity participation, and trip chaining simultaneously.

There are three basic assumptions in the initial SEM models. First, sociodemographics characteristics affect both the activity participation and travel behavior of travelers. Second, the increase of in-home activity participation will reduce

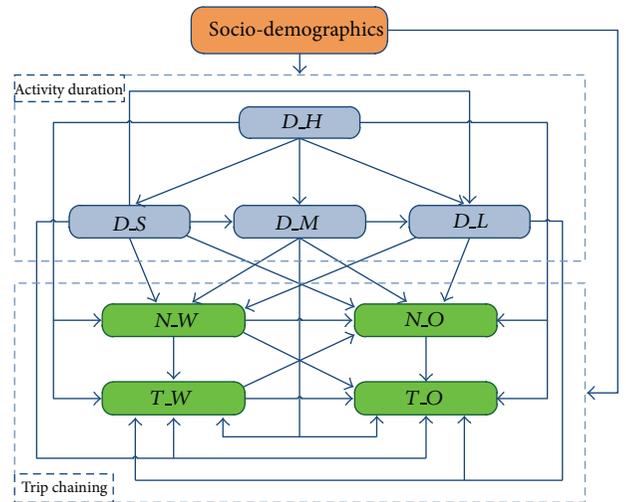


FIGURE 1: Causal structure linking the exogenous variables and endogenous variables.

the time spent on out-of-home activities, the three types of out-of-home activities affect each other mutually. Third,

TABLE 4: Statistical characteristics of the endogenous variables.

Endogenous variables	Low-income group (846 individuals)					Non-low-income group (7534 individuals)				
	Population		Nonzero sample			Population		Nonzero sample		
	Mean	Variance	Mean	Variance	Sample size	Mean	Variance	Mean	Variance	Sample size
$D_H$ (hour)	16.93	4.03	16.93	4.03	846	15.36	3.42	15.34	3.42	7534
$D_S$ (hour)	5.11	4.59	8.86	1.81	488	6.36	4.19	8.67	1.97	5531
$D_M$ (hour)	0.31	0.76	1.17	1.02	225	0.36	0.97	1.25	1.48	2164
$D_L$ (hour)	0.62	1.37	2.18	1.80	238	0.65	1.60	2.55	2.28	1906
$N_W$ (chain)	0.62	0.57	1.08	0.27	488	0.85	0.61	1.16	0.37	5531
$T_W$ (hour)	0.64	0.69	1.12	0.54	488	0.85	0.85	1.15	0.80	5531
$N_O$ (chain)	0.54	0.66	1.22	0.41	373	0.55	0.80	1.42	0.64	2915
$T_O$ (hour)	0.39	0.57	0.88	0.54	373	0.44	0.77	1.14	0.86	2915

TABLE 5: Goodness-of-fit of the two models.

Models	$\chi^2$	DF	$P$	$\chi^2/DF$	RMSEA <sup>1</sup>	GFI <sup>2</sup>	CN <sup>3</sup>
Model A	79.8	88	0.722	0.907	0.000	0.991	1175
Model B	87.7	92	0.607	0.953	0.000	0.999	9912

<sup>1</sup>RMSEA is root mean square error of approximation.

<sup>2</sup>GFI is goodness-of-fit index.

<sup>3</sup>CN is Hoelter's critical  $N$ .

household and individual characteristics not only influence trip chaining behavior directly, but also affect trip chaining indirectly through activity participation of individuals.

The above initial SEM models were estimated by using the software of AMOS 7.0. The maximum likelihood (ML) method was selected as the estimation method, because it converges more rapidly and the results are also easier to interpret compared with the “distribution free” approach (e.g., DWLS) [14]. Generally, the initial model does not perform well; thus, it needs some modification by adding or deleting links according to both their significance which is suggested by the model output and their interpretability. After the modification procedures, we obtained two final models, as shown in Figures 2 and 3.

Table 5 listed goodness-of-fit of the two models. For Model A, which represents the low-income group, the  $\chi^2$  is 79.8 with 88 degrees of freedom, and  $P$  value is 0.722 (greater than 0.05), indicating that the null hypothesis ( $H_0 : \Sigma = \Sigma(\theta)$ ) cannot be rejected. Other measures of fit, such as GFI = 0.991 (that ranges from 0 to 1), and root mean square error of approximation (RMSEA = 0.000) are also found to be acceptable by model fit criteria for structural equation model. Hoelter's critical  $N$  (CN) statistic is found to be 1175 (greater than 200 is considered a good fit), which is the sample size at which value of the fitting function  $F_{ML}$  would lead to the rejection of the null hypothesis,  $H_0$  (i.e.,  $\Sigma = \Sigma(\theta)$ ), at a chosen significance level. Similarly, Model B, which is pertaining to the nonpoor, is also quite satisfactory.

## 6. Model Estimation Results

Tables 6–10 are the estimation results of Model A and Model B. There are three distinct types of relationships that can be obtained from structural equations modeling procedures.

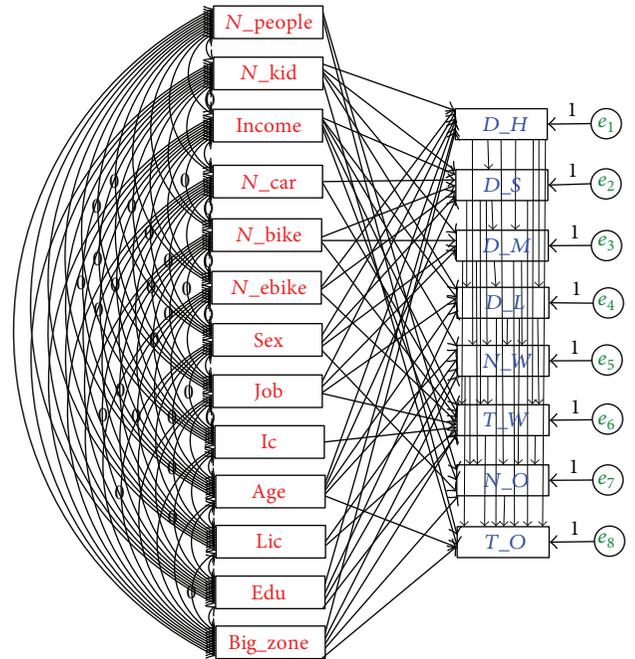


FIGURE 2: SEM path diagram for low-income group.

They are called direct effects, indirect effects, and total effects, respectively. Note that direct and indirect effects may be of different signs, thus having an important implication for the overall total effect. For example, it can be seen in Table 10 (Model A) that the subsistence activity duration ( $D_S$ ) has a negative direct effect ( $-0.121$ ) and positive indirect effect ( $0.149$ ) on the travel time of work chains ( $T_W$ ). Because  $D_S$  has negative direct effects on  $D_M$  and  $D_L$  (e.g.,  $-0.090$  and  $-0.418$ , resp.), both of which have negative direct effects

TABLE 6: Total, direct, and indirect effects of sociodemographics on activity duration and trip chaining in Model A.

	Effects	Big_Zone	N_people	N_kid	Income	N_car	N_bike	N_Ebike	Sex	Job	IC	Age	Lic	Edu
<i>D_H</i>	Total	-0.833	0	0.504	0	0	0	-0.372	1.245	0.257	0	0.554	0	-1.342
	Direct	<b>-0.883</b>	<b>0</b>	<b>0.504</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>-0.372</b>	1.245	0.257	0	0.554	0	-1.342
	Indirect	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>D_S</i>	Total	0.747	0	0.030	-0.270	0.620	0.245	0.731	-1.354	-0.217	0	-0.922	0	1.136
	Direct	<b>0</b>	<b>0</b>	<b>0.457</b>	<b>-0.270</b>	<b>0.620</b>	<b>0.245</b>	<b>0.416</b>	-0.300	0	0	-0.453	0	0
	Indirect	0.747	0	-0.427	0	0	0	0.315	-1.054	-0.217	0	-0.469	0	1.136
<i>D_M</i>	Total	-0.049	0	0.084	0.024	-0.056	0.055	-0.058	0.305	0.037	0	0.072	0	-0.075
	Direct	<b>0</b>	<b>0</b>	<b>0.097</b>	<b>0</b>	<b>0</b>	<b>0.077</b>	<b>0</b>	0.209	0.023	0	0	0	0
	Indirect	-0.049	0	-0.013	0.024	-0.056	-0.022	-0.058	0.096	0.014	0	-0.072	0	-0.075
<i>D_L</i>	Total	0.003	0	-0.143	0.045	-0.216	-0.144	-0.144	-0.060	0.033	0	0.229	0	-0.197
	Direct	<b>0</b>	<b>0</b>	<b>0.093</b>	<b>-0.050</b>	<b>0</b>	<b>0</b>	<b>0</b>	0	0.051	0	0.073	0	-0.202
	Indirect	0.003	0	-0.236	0.094	-0.216	-0.144	-0.144	-0.060	-0.018	0	0.156	0	0.005
<i>N_W</i>	Total	0.133	0	-0.037	0.021	0.042	0.005	0.052	-0.131	-0.031	0	-0.119	-0.087	0.099
	Direct	<b>0.077</b>	<b>0</b>	<b>0</b>	<b>0.028</b>	<b>0</b>	<b>0</b>	<b>0</b>	0	0	0	-0.037	-0.087	-0.032
	Indirect	0.056	0	-0.037	-0.007	0.042	0.005	0.052	-0.131	-0.031	0	-0.082	0	0.131
<i>T_W</i>	Total	0.197	0	-0.081	0.065	-0.069	-0.046	0.017	-0.153	-0.037	-0.068	-0.101	0.013	0.263
	Direct	<b>0.090</b>	<b>0</b>	<b>0</b>	<b>0.045</b>	<b>-0.086</b>	<b>-0.041</b>	<b>-0.029</b>	0	0.008	-0.068	0	0.042	0.080
	Indirect	0.107	0	-0.081	0.020	0.017	-0.005	0.045	-0.153	-0.045	0	-0.101	-0.029	0.184
<i>N_O</i>	Total	-0.038	0.020	0.032	0.031	-0.065	-0.007	-0.073	0.136	0.038	0.008	0.128	0.023	-0.155
	Direct	<b>0.059</b>	<b>0.020</b>	<b>0</b>	<b>0.027</b>	<b>0</b>	<b>0</b>	<b>0</b>	-0.038	0	0	0	0	0
	Indirect	-0.097	0	0.032	0.004	-0.065	-0.007	-0.073	0.174	0.038	0.008	0.128	0.023	-0.155
<i>T_O</i>	Total	-0.049	-0.002	-0.023	0.011	-0.051	-0.006	-0.059	0.107	0.023	0.021	0.117	0.018	-0.112
	Direct	<b>0.018</b>	<b>-0.011</b>	<b>-0.022</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	0	0	0	0.009	0	0
	Indirect	-0.067	0.009	-0.001	0.011	-0.051	-0.006	-0.059	0.107	0.023	0.021	0.108	0.018	-0.112

(-0.312, -0.281) on *T\_W*. According to the effect analysis theory, the indirect effects of *D\_S* on *T\_W* can be computed as  $(-0.090) \times (-0.312) + (-0.418) \times (-0.281) = 0.149$ . Therefore, the total effect (0.028) of *D\_S* on *T\_W* is the algebraic sum of direct effect (-0.121) and indirect effect (0.149).

It can be found that strong relationship exists among the sociodemographics, activity participation, and travel behavior, both for the poor and the nonpoor. In the following, we will examine the effects in detail from 4 aspects: effects of sociodemographics on activity duration and trip chaining, effects of activity durations on each other, effects of trip-chaining on trip chaining, and effects of activity duration on trip chaining behavior.

**6.1. Effects of Sociodemographics on Activity Duration and Trip Chaining.** From Tables 6 and 7, we can see that, in both groups, some sociodemographics significantly affect all four types of activities and four trip chaining variables. The household and individual characteristics that are systematically important in explaining variations in activity participation and travel behavior include house location, income, number of preschool children, age, gender, and educational level.

Combining the path diagram in Figures 2 and 3, it can also be found that household characteristics have more influence on the activity participation of low-income population (12 routes from household characteristics to activity durations and 11 routes from individual characteristics to activity durations), while individual characteristics have more influence

on the activity participation of the nonpoor (7 routes from household characteristics to activity durations and 17 routes from individual characteristics to activity durations). In addition, sociodemographics have more direct influence (22 routes) on the trip-chaining in the low-income group than that of the nonpoor (17 routes).

Specifically, the number of preschool children significantly affects the activity duration of the low-income group, but it has no effects on that of the non-low-income group. On the contrary, the IC factor does not influence low-income populations' activity duration at all but has significant effects on non-income populations.

**6.2. Effects of Activity Duration on Activity Duration.** From Table 8, it can be found that interaction effects among 4 activity durations follow the same framework both in Model A and Model B. *D\_H* has negative direct effects on the duration of out-of-home activities, *D\_M* has negative effects on *D\_M* and *D\_L*, and *D\_M* has negative effects on *D\_L*.

However, the values of effects are not quite similar in the two models. For example, the absolute values of effects of *D\_H* on other activity durations in Model A are all smaller than those in model B, while the effects of *D\_M* on *D\_L* in Model A are larger than their counterparts in model B, which indicates that the trade-off among the 4 type activities is differently in two groups. It can be interpreted that low-income population spend more time at home and have lower value of time due to their inferior social status and limited social network.

TABLE 7: Total, direct, and indirect effects of sociodemographics on activity duration and trip chains in Model B.

	Effects	Big_Zone	N_people	N_kid	Income	N_car	N_bike	N_Ebike	Sex	Job	IC	Age	Lic	Edu
<i>D.H</i>	Total	0	0	0	0.103	0	0	0	-1.158	0	-0.285	0.19	0	0
	Direct	0	0	0	0.103	0	0	0	-1.158	0	-0.285	0.19	0	0
	Indirect	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>D.S</i>	Total	0.424	-0.119	0	-0.273	0.159	0	0	1.365	0.054	0.547	-0.189	0	0
	Direct	0.424	-0.119	0	-0.171	0.159	0	0	0.213	0.054	0.263	0	0	0
	Indirect	0	0	0	-0.102	0	0	0	1.152	0	0.283	-0.189	0	0
<i>D.M</i>	Total	-0.009	0.017	0	0.093	-0.022	0.017	0	-0.247	-0.029	-0.067	-0.003	0	0
	Direct	0.05	0	0	0.058	0	0.017	0	-0.097	-0.021	0	-0.023	0	0
	Indirect	-0.059	0.017	0	0.035	-0.022	0	0	-0.15	-0.008	-0.067	0.02	0	0
<i>D.L</i>	Total	-0.201	0.096	0	0.149	-0.068	-0.007	0	-0.226	-0.053	-0.131	-0.016	0	-0.13
	Direct	0	0.045	0	0.092	0	0	0	-0.097	-0.038	0	-0.038	0	-0.13
	Indirect	-0.201	0.051	0	0.057	-0.068	-0.007	0	-0.129	-0.015	-0.131	0.022	0	0
<i>N.W</i>	Total	0.127	-0.018	-0.031	0.012	0.026	-0.003	0	0.138	-0.005	0.078	-0.016	-0.031	-0.016
	Direct	0.105	0	-0.031	0	0	0	0	0	-0.021	0.038	-0.014	-0.031	0.035
	Indirect	0.022	-0.018	0	0.012	0.026	-0.003	0	0.138	0.016	0.04	-0.002	0	-0.05
<i>T.W</i>	Total	-0.044	-0.019	-0.009	0.002	0.049	0.005	-0.008	0.243	0.014	0.059	-0.014	-0.009	-0.037
	Direct	-0.069	0	0	0	0	0.009	-0.008	0	0	0	0	0	0.02
	Indirect	0.025	-0.019	-0.009	0.002	0.049	-0.004	0	0.243	0.014	0.059	-0.014	-0.009	-0.057
<i>N.O</i>	Total	-0.052	0.034	0.005	-0.027	-0.039	0.005	0.007	-0.217	-0.024	-0.081	-0.01	0.005	0.096
	Direct	<b>0.019</b>	0	0	0	0	0	0.006	0	0	0	-0.014	0	0.017
	Indirect	<b>-0.071</b>	0.034	0.005	-0.027	-0.039	0.005	0	-0.217	-0.024	-0.081	0.004	0.005	0.079
<i>T.O</i>	Total	-0.099	0.034	0.007	-0.03	-0.033	0.002	0.006	-0.259	-0.02	-0.084	-0.005	0.007	0.077
	Direct	-0.036	0	0	0	0	0	0	-0.069	0	0	0	0	0
	Indirect	-0.063	0.034	0.007	-0.03	-0.033	0.002	0.006	-0.19	-0.02	-0.084	-0.005	0.007	0.077

TABLE 8: Total, direct, and indirect effects of activity duration on activity duration.

	Effects	Model A				Model B			
		<i>D.H</i>	<i>D.S</i>	<i>D.M</i>	<i>D.L</i>	<i>D.H</i>	<i>D.S</i>	<i>D.M</i>	<i>D.L</i>
<i>D.H</i>	Total	0	0	0	0	0	0	0	0
	Direct	0	0	0	0	0	0	0	0
	Indirect	0	0	0	0	0	0	0	0
<i>D.S</i>	Total	-0.847	0	0	0	-0.995	0	0	0
	Direct	-0.847	0	0	0	-0.995	0	0	0
	Indirect	0	0	0	0	0	0	0	0
<i>D.M</i>	Total	0.056	-0.090	0	0	0.104	-0.141	0	0
	Direct	-0.021	-0.090	0	0	-0.036	-0.141	0	0
	Indirect	0.076	0	0	0	0.140	0	0	0
<i>D.L</i>	Total	-0.004	-0.349	-0.765	0	0.066	-0.427	-0.400	0
	Direct	-0.315	-0.418	-0.765	0	-0.374	-0.484	-0.400	0
	Indirect	0.311	-0.069	0	0	0.440	0.056	0	0

6.3. *Effects of Trip Chaining on Trip Chaining.* According to Table 9, the effects of trip-chaining characteristics on each other also follow similar frameworks in both models. Specifically, *N.W* has positive effects on *T.W* and negative effects on both *N.O* and *T.O*, *T.W* has negative effects on *N.O* and *T.O*, and *N.O* has negative effects on *T.O*, which indicates that there are strong relationships and trade-offs between work chains and non-work chains.

Note that the absolute value of effects of work chains on non-work chains of the poor is larger than that of the

nonpoor. It can be explained that low-income residents have less freedom to participate in different types of activities other than work due to their economic status.

6.4. *Effects of Activity Duration on Trip Chaining.* The estimation results in Table 10 show that, both for the poor and nonpoor, travel is derived from activity participation. Activity duration also affects trip chaining behavior besides sociodemographics. For example, we find that number of work chains (*N.W*) and the travel time of work chains (*T.W*)

TABLE 9: Total, direct, and indirect effects of trip chaining on trip chaining.

Effects	Model A				Model B			
	<i>N_W</i>	<i>T_W</i>	<i>N_O</i>	<i>T_O</i>	<i>N_W</i>	<i>T_W</i>	<i>N_O</i>	<i>T_O</i>
<i>N_W</i>	Total	0	0	0	0	0	0	0
	Direct	0	0	0	0	0	0	0
	Indirect	0	0	0	0	0	0	0
<i>T_W</i>	Total	0.334	0	0	0	0.308	0	0
	Direct	0.334	0	0	0	0.308	0	0
	Indirect	0	0	0	0	0	0	0
<i>N_O</i>	Total	-0.318	-0.113	0	0	-0.167	-0.056	0
	Direct	-0.281	-0.113	0	0	-0.150	-0.056	0
	Indirect	-0.037	0	0	0	-0.017	0	0
<i>T_O</i>	Total	-0.349	-0.302	0.433	0	-0.237	-0.274	0.600
	Direct	-0.127	-0.253	0.433	0	-0.063	-0.241	0.600
	Indirect	-0.222	-0.049	0	0	-0.174	-0.033	0

TABLE 10: Total, direct, and indirect effects of activity duration on trip chaining.

Effects	Model A				Model B				
	<i>D_H</i>	<i>D_S</i>	<i>D_M</i>	<i>D_L</i>	<i>D_H</i>	<i>D_S</i>	<i>D_M</i>	<i>D_L</i>	
<i>N_W</i>	Total	-0.063	0.067	-0.151	-0.226	-0.072	0.076	-0.197	-0.201
	Direct	<b>-0.081</b>	<b>-0.041</b>	<b>-0.324</b>	<b>-0.226</b>	<b>-0.079</b>	<b>-0.049</b>	<b>-0.278</b>	<b>-0.201</b>
	Indirect	0.018	0.108	0.173	0	0.007	0.125	0.081	0
<i>T_W</i>	Total	-0.091	0.028	-0.148	-0.356	-0.154	0.014	-0.256	-0.376
	Direct	<b>-0.157</b>	<b>-0.121</b>	<b>-0.312</b>	<b>-0.281</b>	<b>-0.266</b>	<b>-0.189</b>	<b>-0.320</b>	<b>-0.314</b>
	Indirect	0.065	0.149	0.164	-0.076	0.111	0.203	0.064	-0.062
<i>N_O</i>	Total	0.071	-0.120	0.237	0.300	0.105	-0.172	0.311	0.303
	Direct	0.025	0	0.329	0.197	0.030	0	0.368	0.252
	Indirect	0.046	-0.120	-0.091	0.103	0.075	-0.172	-0.057	0.051
<i>T_O</i>	Total	0.043	-0.124	0.159	0.207	0.086	-0.191	0.259	0.258
	Direct	-0.080	-0.074	-0.031	-0.041	-0.108	-0.093	-0.013	-0.027
	Indirect	0.123	-0.050	-0.191	0.249	0.194	-0.098	0.272	0.285

are directly affected by the 4 categories of activities. We also find that the *N\_W* increases as the *D\_M* or *D\_L* decreases, while *N\_O* increases as *D\_M* or *D\_L* increases.

From Table 10, we can also find that, in different models, the effects between the same variables are different. For example, the absolute value of *total effect* of *D\_M* on *N\_W* is smaller in the low-income group than in the non-low-income group; however, the corresponding absolute values of *direct effect* and *indirect effect* are larger in Model A than that in Model B.

In general, the absolute values of effects in Model B are greater than their counterparts in Model A, which indicates that activity participation has larger effects on the trip chaining characteristics in the group of non-low-income population than that of the poor. This scenario is possibly because the sample size of the low-income group (846) is much less than that of the non-low-income group (7534); thus, on the whole, the causal relationship of activity duration and trip chaining behavior revealed in Model B is stronger than that in Model A.

## 7. Conclusions

This paper focuses on the activity-trip chaining behavior of urban low-income populations in developing countries. Using the data of residents travel survey of Nanjing City (2009) and a specific travel survey of low-income residents of Nanjing City (2010), we proposes two structural equation models to investigate the relationships among sociodemographics, activity participation, and travel behavior of both low-income populations and non-low-income populations of Nanjing City. Based on the model outputs, we analyzed four categories of effects of the two groups. The general findings can be summarized as follows.

First, on average, the duration of out-of-home activities taken by the low-income populations is less than that of the non-low-income populations, and the less trip chains and less total travel time indicate that low-income populations generally do less out-of-home activities.

Second, the relationships among sociodemographics, activity duration, and trip chaining of both groups can be captured by the proposed SEM models, and most of the

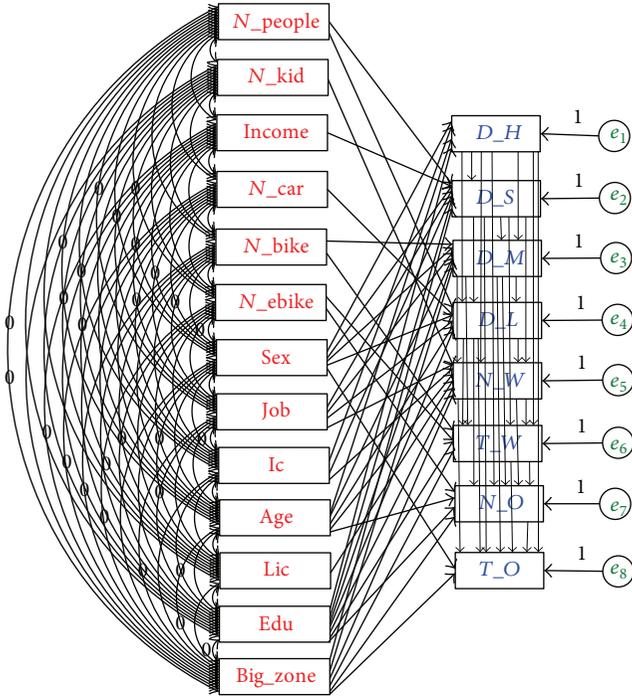


FIGURE 3: SEM path diagram for non-low-income group.

estimated effects are quite similar to those reported in the literature.

Third, both the structural equation models follow the same modeling framework. Therefore, the activity-trip chaining behavior of both the low-income populations and non-low-income populations shares some similarities. For example, sociodemographics, especially household income; residential locations; age; and gender significantly affects the activity-trip chaining behavior of both the poor and the nonpoor.

Finally, low-income populations have some unique characteristics on the activity-travel behavior, which are different from those of the non-low-income populations. For instance, household characteristics have more influence on the activity participation of low-income population; the trade-off among the four type activities is differently in two groups; the effects of work chains on non-work chains of the poor are larger than those of the nonpoor; in general, activity participation has greater effects on trip chaining in the group of non-low-income residents than that of low-income residents.

Based on these findings of travel behavior characteristics of urban low-income populations in developing countries, the following policies are suggested for the government and transportation agencies.

- (1) Adopt transit-oriented transportation planning strategy, such as adding new shuttle buses from low-income population concentrated areas to metro stations and opening new bus lines across low-income neighborhoods step by step.
- (2) In order to reduce the monetary cost of low-income residents, the government can either subsidize them

directly to improve social equity or introduce two or more bus operating companies to break the monopoly, so as to improve the level of bus service and reduce the bus fares.

- (3) In the long-term planning, the city should transform from single center pattern to polycentric development pattern. Meanwhile, the government should consider hybrid land use and create more job opportunities near the residential area of low-income populations, such that low-income residents in the urban fringe will not waste too much time on their trip chains.
- (4) Provide more vocational training for low-income adults and improve their ability of earning money. In addition, guarantee the next generation of low-income residents can receive high quality education and help them climb higher along the social ladder. These policies can change their inferior position of travel fundamentally.

This research offers promising insights into the activity-travel behavior of the poor and extends the need to crafting effective transportation policies specifically for the urban low-income populations in developing countries. However, this research can be extended in terms of the following aspects: (a) conduct specific studies on the trading-off relationships between in-home and out-of-home activities; (b) study the interactions between activity participation and travel chaining behavior on two or more successive days; (c) consider the household level activity-travel behavior characteristics instead of individual level; (d) adopt the proposed SEM model to other cities in developing countries. It is hoped that these issues and others can be addressed in the future.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research is supported by the National Natural Science Foundation of China (NSFC, no. 51078085, 51178109, 51178110, 51378119), Graduate Innovation Project of Jiangsu Province (No. CXZZ12\_0113), and the Fundamental Research Funds for the Central Universities, China. The authors would like to express their appreciation towards Nanjing Institute of City & Transport Planning Co., Ltd., in particular, for the valuable assistance in obtaining and interpreting the data used for these models.

## References

- [1] G. Giuliano, H. Hu, and K. Lee, "The role of public transit in the mobility of low income households," Final Report, Metrans Transportation Center, Los Angeles, Calif, USA, 2001, <http://www.americandreamcoalition.org/automobility/transitfor-poor.pdf>.
- [2] G. Giuliano, "Low income, public transit, and mobility," *Transportation Research Record*, no. 1927, pp. 63–70, 2005.

- [3] E. Blumenberg and P. Haas, "The travel behavior and needs of the poor: a study of welfare recipients in Fresno County," Publication FHWA-CA-OR-2001-23, FHWA, U.S. Department of Transportation, 2001.
- [4] K. Clifton, "Examining travel choices of low-income populations—issues, methods, and new approaches," in *Proceedings of the 10th International Conference on Travel Behavior Research*, Lucerne, Switzerland, August 2003.
- [5] N. McDonald, S. Librera, and E. Deakin, "Free transit for low-income youth: experience in San Francisco Bay area, California," *Transportation Research Record*, no. 1887, pp. 153–160, 2004.
- [6] R. Behrens, "Understanding travel needs of the poor: Towards improved travel analysis practices in South Africa," *Transport Reviews*, vol. 24, no. 3, pp. 317–336, 2004.
- [7] S. Srinivasan and P. Rogers, "Travel behavior of low-income residents: studying two contrasting locations in the city of Chennai, India," *Journal of Transport Geography*, vol. 13, no. 3, pp. 265–274, 2005.
- [8] P. Thakuriah, P. S. Sriraj, S. Sööt, and Y. Liao, "Determinants of perceived importance of targeted transportation services for low-income riders," *Transportation Research Record*, no. 1986, pp. 145–153, 2006.
- [9] J. Taylor, M. Barnard, H. Neil, and C. Creegan, *The Travel Choices and Needs of Low Income Households: The Role of the Car*, The National Centre for Social Research, London, UK, 2009, <http://trid.trb.org/view.aspx?id=886473>.
- [10] S. Gao and R. A. Johnston, "Public versus private mobility for low-income households: transit improvements versus increased car ownership in the sacramento, California, region," *Transportation Research Record*, no. 2125, pp. 9–15, 2009.
- [11] T. F. Golob, "Structural equation modeling for travel behavior research," *Transportation Research B: Methodological*, vol. 37, no. 1, pp. 1–25, 2003.
- [12] R. Kitamura, J. P. Robinson, T. F. Golob, M. A. Bradley, J. Leonard, and T. van der Hoorn, "A comparative analysis of time use data in the Netherlands and California," in *Proceedings of the 20th PTRC Summer Annual Meeting: Transportation Planning Methods*, pp. 127–138, 1992.
- [13] X. Lu and E. I. Pas, "Socio-demographics, activity participation and travel behavior," *Transportation Research A: Policy and Practice*, vol. 33, no. 1, pp. 1–18, 1999.
- [14] T. F. Golob, "A simultaneous model of household activity participation and trip chain generation," *Transportation Research B: Methodological*, vol. 34, no. 5, pp. 355–376, 2000.
- [15] A. R. Kuppam and R. M. Pendyala, "A structural equations analysis of commuters' activity and travel patterns," *Transportation*, vol. 28, no. 1, pp. 33–54, 2001.
- [16] J.-H. Chung and Y. Ahn, "Structural equation models of day-to-day activity participation and travel behavior in a developing country," *Transportation Research Record*, no. 1807, pp. 109–118, 2002.
- [17] M. Yang, W. Wang, X. Chen, T. Wan, and R. Xu, "Empirical analysis of: commute trip chaining case study of Shangyu, China," *Transportation Research Record*, no. 2038, pp. 139–147, 2007.
- [18] S. S. V. Subbarao and K. V. Krishna Rao, "Trip chaining behavior in developing countries: a study of Mumbai Metropolitan Region, India," *European Transport*, paper 3, no. 53, pp. 1–30, 2013.
- [19] M. Yang, W. Wang, G. Ren, R. Fan, B. Qi, and X. Chen, "Structural equation model to analyze sociodemographics, activity participation, and trip chaining between household heads: survey of Shangyu, China," *Transportation Research Record*, no. 2157, pp. 38–45, 2010.

## Research Article

# Rough Set Approach for Group Evacuation Behavior Analysis in Passenger Transport Hub Area

Peng Chen,<sup>1</sup> Xiaofeng Ma,<sup>2</sup> and Shunying Zhu<sup>1</sup>

<sup>1</sup> School of Transportation, Wuhan University of Technology, Wuhan, Hubei 430063, China

<sup>2</sup> Engineering Research Center for Transportation Safety (Ministry of Education), Wuhan University of Technology, Wuhan, Hubei 430063, China

Correspondence should be addressed to Peng Chen; seupc@163.com

Received 3 November 2013; Revised 27 February 2014; Accepted 2 March 2014; Published 17 April 2014

Academic Editor: Huimin Niu

Copyright © 2014 Peng Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Evacuation behavior analysis is deemed to be one aspect of evacuation planning. However, existing studies have not discussed group evacuation decision-making in the face of disagreement among decision makers. In this paper, rough set theory is applied to analyze group evacuation decision-making in passenger transport hub area with various groups including kin, lover, friend, colleague, and classmate. In the approach, improved tabu search-based attribute reduction is proposed to find the minimal subset of attributes required to fully describe the information of group evacuation decision-making, and value reduction algorithm based on knowledge granulation is used to generate rules of group evacuation decision-making. Cross-validation procedure is adopted to estimate the performance of rough set theory. Experimental results indicate that rough set theory has favorable performance. Thus, the proposed approach provides a new way for evacuation behavior analysis.

## 1. Introduction

Urban passenger transport hub works as the joint of inter-modal transit and the distribution center of massive passenger flow. In such densely populated area, small scale emergencies can result in severe consequences and should not be overlooked. It is vital to evacuate people from the affected area promptly. Thus, evacuation planning is very crucial. Evacuation behavior analysis is deemed to be one aspect of evacuation planning. Understanding emergency evacuation behavior would help better emergency evacuation planning.

Over the last few decades, considerable research has focused on evacuation behavior analysis related to hurricane evacuation [1–6] and building fire evacuation [7, 8]. The majority of existing studies focused on individual and household evacuation decision-making and behavioral responses. However, it has not been discussed that how households arrive at a decision when decision-making of household members is inconsistent. As for emergency evacuation in

passenger transport hub area, there are usually a large number of small groups with various relationships, including kin, lover, friend, colleague, and classmate. It is necessary to address the problem about group evacuation decision-making in the face of disagreement among decision makers.

By far, several methods have been put forward for helping us understand evacuation decision-making, including contingency table analysis [8], artificial neural network [7], and logistic regression analysis [4, 6, 9, 10]. Among the above methods, contingency table analysis is only used to determine whether dependence exists between evacuation decision-making and related factors for a given significant level. Artificial neural network can establish the mapping function from explanatory variables to evacuation decision-making by using the multilayer perceptron, but it can only provide implied knowledge of evacuation decision-making. Logistic regression model is adopted to describe the relationship between explanatory variables and the probability of outcome of evacuation decision-making, but it needs to create the complex function. Thus rough set theory [11–13]

TABLE 1: Chi-square test for group decision-making with several factors.

	Factors							
	Age	Gender	Education	Temperament	Experience	Number of luggage	Familiar with route	Group relationship
Chi-square	24.604	22.489	25.258	26.315	31.02	12.804	10.303	32.467
P-value	0.017	0.004	0.014	0.01	0.002	0.383	0.85	0.001

is proposed to analyze group evacuation decision-making under emergency evacuation in passenger transport hub area. The approach adopted in this paper does not require the establishment of the function and can generate rules expressed in the form of if-then statements, which make knowledge have a clear meaning.

Rough set theory can remove redundant information through the reduction and extract decision rules from a large number of original data on the premise of the maintenance of the same classification ability. Knowledge reduction, including attribute reduction and value reduction, is one of the core issues in rough set theory. On the one hand, attribute reduction tends to reduce the complexity and cost of decision process and promote higher rule quality. In order to compute useful reduction of information systems, many researchers have developed some efficient algorithms based on computational intelligence tools of genetic algorithm [14], ant colony optimization [15], simulated annealing [15], particle swarm optimization [16], tabu search [17], and so on. On the other hand, value reduction is aimed at elimination of redundant attribute values in each rule and simplification of rules set while keeping the classification ability of rules set. Several algorithms have been proposed for value reduction, such as simplification rule algorithm [18], discernibility matrix-based algorithm [19], and heuristic algorithm based on mutual information [20]. In this paper, we propose improved tabu search-based attribute reduction (ITSAR) to find the minimal subset of attributes required to fully describe the information of group evacuation decision-making. Different from the existing research [17], we measure solution quality based on knowledge granulation because it can overcome the shortcomings of dependency degree [21] and use dynamic tabu tenure because it has better performance than fixed tabu tenure [22–24]. We also propose a heuristic algorithm based on knowledge granulation for value reduction, which is used to generate the rules of group evacuation decision-making.

The remainder of this paper is organized as follows. The next section describes evacuation behavior survey in Wuchang Railway Station area for the preparation of data set used in this study. In the following section, rough set theory is introduced, including related concepts, the algorithms for attribution reduction and value reduction, and evaluation of the approach. Section four presents the application of rough set theory on group evacuation decision-making and compares the proposed method with other methods in performance. Finally, we conclude the paper with a summary and outlook for further research.

## 2. Evacuation Behavior Survey

A survey was conducted about emergency evacuation behavior in Wuchang Railway Station area, with the hypothetical event of the toxic gas attack. The questionnaire was designed to collect the following information related to human behavior: (1) personal information including age, gender, education, temperament, and the number of luggage; (2) familiar with the route or not familiar with the route; (3) past experience; (4) the number of group members and group relationship; (5) human behavioral response including first action, evacuation route choice, group evacuation decision-making, and so on. Among the above information, the question for past experience is “Did you ever experience gas attack or participate in safety training,” and structured answer is “(1) Never experience gas/training experience/knowledge, (2) Have gas experience/training experience/knowledge.” The structured answer for temperament is “(1) Choleric (You are a strong-willed individual who makes decisions quickly and decisively.), (2) Sanguine (You are affectionate, enjoy social activities, and make friends easily.), (3) Phlegmatic (You are dependable, polite, and even-tempered.), and (4) Melancholic (Time alone is vital for this reflective, introspective temperament.)”.

A total of 952 interviews were performed and 909 valid replies were collected. There were 523 (57.5%) valid replies coming from groups and 386 (42.5%) from the single passenger. This paper focuses on the analysis of group evacuation decision-making in the face of disagreement among decision makers. In order to select the attributes influencing group evacuation decision-making, contingency table analysis was performed to test the correlation between group decision-making and the characteristics of individual and group by utilizing statistical analysis software SPSS 19.0. As shown in Table 1, the results of  $\chi^2$  (Chi-square) test at a significance value ( $\alpha = 0.05$ ) indicate that there exists a significant relationship between group decision-making and several factors as follows: age, gender, education, temperament, experience, and group relationship. Based on group decision-making and the above factors, the attribute set and attribute value set are listed in Table 2.

## 3. Rough Set Theory

This section introduces rough set theory. Some basic notions are introduced in Section 3.1. Then the algorithms for attribute reduction and value reduction are developed in Sections 3.2 and 3.3, respectively. Section 3.4 explains evaluation of the approach.

TABLE 2: The attribute set and attribute value set.

Class	Attribute set	Attribute value set
Condition attribute	Age (C1)	① <18, ② 18–35, ③ >35
	Gender (C2)	① Female, ② male
	Education (C3)	① Senior high school or below (SHS), ② junior college or above (JC)
	Temperament (C4)	① Choleric, ② sanguine, ③ phlegmatic, and ④ melancholic
	Experience (C5)	① Never experience gas/training experience/knowledge (NE), ② have gas experience/training experience/knowledge (HE)
	Group relationship (C6)	① Kin, ② lover, ③ friend, colleague, or classmate (FC)
Decision attribute	Group decision-making (D)	① The minority is subordinate to the majority (MSM); ② choose the route approved by self (ABS); ③ choose the route approved by the one familiar with the route (ABF); ④ choose the route approved by the one doing things reasonably (ABR)

**3.1. Preliminaries.** In this section, some preliminary concepts such as indiscernibility, knowledge granulation, attribute reduction, and value reduction are briefly presented.

**3.1.1. Indiscernibility.** Let  $S = (U, A)$  be an information system, where  $U$ , called universe, is a nonempty set of finite objects;  $A$  is a nonempty finite set of attributes such that  $a : U \rightarrow V_a$  for every  $a \in A$ ;  $V_a$  is the value set of  $a$ . In a decision system with a set of decision attributes,  $A = C \cup D$ , where  $C$  is the set of condition attributes and  $D$  is the set of decision attributes. Such an information system also is called a decision table.

For an attribute set  $P \subseteq A$ , there is an associated indiscernibility relation  $\text{IND}(P)$ :

$$\text{IND}(P) = \{(x, y) \in U^2 \mid \forall a \in P, a(x) = a(y)\}. \quad (1)$$

$U/P$  denotes the partition of  $U$  generated by  $\text{IND}(P)$ . If  $(x, y) \in \text{IND}(P)$ , then  $x$  and  $y$  are indiscernible by attributes from  $P$ .  $[x]_P$  denotes the equivalence classes of the  $P$ -indiscernibility relation. The indiscernibility relation is the mathematical basis of rough set theory.

### 3.1.2. Knowledge Granulation

**Definition 1.** Let  $S = (U, A)$  be an information system and  $U/A = \{X_1, X_2, \dots, X_m\}$ ; then knowledge granulation of  $A$  is given by

$$\text{GK}(A) = \frac{1}{|U|^2} \sum_{i=1}^m |X_i|^2, \quad (2)$$

where the symbol  $|\cdot|$  means the cardinality of a set.

**Definition 2.** Let  $S = (U, C \cup D)$  be a decision table, where  $U$  is the universe,  $C$  is the set of conditional attributes and  $D$  is the decision attribute,  $B \subseteq C$ ; then the relative partition granularity of  $B$  relative to  $D$  is defined by Feng et al. [25]:

$$\text{GK}(D \mid B) = \text{GK}(B) - \text{GK}(B \cup D). \quad (3)$$

The value of  $\text{GK}(D \mid B)$  can be used to measure the classification ability of  $B$  relative to  $D$ ; that is, the larger the value of  $\text{GK}(D \mid B)$ , the weaker the classification ability of  $B$  relative to  $D$ .

**Definition 3.** Let  $S = (U, C \cup D)$  be a decision table, where  $U$  is the universe,  $C$  is the set of conditional attributes, and  $D$  is the decision attribute,  $a \in C$ ; then the significance of attribute  $a$  in  $C$  relative to  $D$  is defined by

$$\text{Sig}(a, C, D) = \text{GK}(D \mid C - \{a\}) - \text{GK}(D \mid C). \quad (4)$$

**3.1.3. Attribute Reduction.** Attribute reduction in rough set theory can preserve the information content while reducing the number of attributes involved. Based on relative partition granularity, a relative reduct can be defined by the following definition.

**Definition 4.** Let  $S = (U, C \cup D)$  be a decision table, where  $U$  is the universe,  $C$  is the set of conditional attributes, and  $D$  is the decision attribute,  $P \subseteq C$ ; if  $\text{GK}(D \mid P) = \text{GK}(D \mid C)$  and  $\text{Sig}(a, P, D) > 0$ ,  $a \in P$ , then  $P$  is said to be a relative reduct of  $C$  relative to  $D$ .

In particular, a relative reduct with minimal cardinality is called minimal reduct. The goal of attribute reduction is to find a minimal reduct.

**3.1.4. Value Reduction.** The process by which the maximum number of condition attribute values is removed without losing essential information is called value reduction. After value reduction, rules can be generated by associating the condition attribute values with the corresponding decision class value.

**Definition 5.** Let  $S = (U, C \cup D)$  be a decision table, and let  $X_i \in U \mid C$ ,  $Y_j \in U \mid D$ , and  $X_i \cap Y_j \neq \emptyset$ . By  $\text{des}(X_i)$  and  $\text{des}(Y_j)$ , we denote the descriptions of the equivalence classes  $X_i$  and  $Y_j$  in the decision table  $S$ . A decision rule is formally defined as

$$Z_{ij} : \text{des}(X_i) \longrightarrow \text{des}(Y_j). \quad (5)$$

*Definition 6.* The confidence of decision rule  $Z_{ij}$  is defined as

$$\text{Con}(Z_{ij}) = \frac{|X_i \cap Y_j|}{|X_i|}. \quad (6)$$

For a certain rule,  $\text{Con}(Z_{ij}) = 1$ , whereas an uncertain rule,  $0 < \text{Con}(Z_{ij}) < 1$ .

**3.2. Improved Tabu Search for Attribute Reduction.** In this section, improved tabu search-based attribute reduction (ITSAR) is proposed to find a minimal reduct of group evacuation decision-making. First we introduce the main idea of tabu search, then describe the components of ITSAR, and finally give the ITSAR scheme.

**3.2.1. Main Idea of Tabu Search.** Tabu search (TS) is a metaheuristic optimization method originally proposed by Glover [26]. TS has been successfully applied in various fields [23, 24, 27]. The main ideas are to avoid recently visited parts of the solution space and to guide the search towards new and promising areas. Nonimproving moves are allowed to escape from local optima, and attributes of recently performed moves are declared tabu or forbidden for a number of iterations to avoid cycling. During the search, the algorithm maintains short-term and long-term memory structures. The short-term memory is built to keep the recency by constructing Tabu List (TL). The long-term memory is utilized to record solutions of special characters like elite and frequently visited solutions.

**3.2.2. Solution Representation.** ITSAR uses a binary representation for solutions (attribute subsets). Therefore, a trial solution  $x$  is a 0-1 vector with dimension equal to the number of condition attributes  $|C|$ . If a component  $x_i$  of  $x$ ,  $i = 1, \dots, |C|$ , has the value 1, then the  $i$ th attribute is contained in the attribute subset represented by the trial solution  $x$ . Otherwise, the solution  $x$  does not contain the  $i$ th attribute.

**3.2.3. Solution Quality Measure.**  $\text{GK}(D \mid x)$  means the relative partition granularity of solution  $x$  relative to decision attribute  $D$ . Comparing two solutions  $x$  and  $x'$ , we say  $x$  is better than  $x'$  if one of the following conditions holds:

$$\begin{aligned} \text{GK}(D \mid x) &< \text{GK}(D \mid x'), \\ \sum_i x_i &< \sum_i x'_i \quad \text{if } \text{GK}(D \mid x) = \text{GK}(D \mid x'). \end{aligned} \quad (7)$$

**3.2.4. Tabu List.** The role of Tabu List (TL) is to avoid being trapped in local optima. The first and second positions in TL are permanently reserved for two special solutions: solution of all ones (i.e., all attributes are considered), and solution of all zeroes (i.e., all attributes are discarded). The remaining positions in TL are used to save the most recently visited solutions. To improve search performance, dynamic selection strategies of tabu tenure are as follows.

The range of tabu tenure  $t$  is defined by parameters  $t_{\min}$  and  $t_{\max}$ . The initial tabu tenure is set equal to

$\text{round}((t_{\min} + t_{\max})/2)$ . In the process of the implementation of diversification strategy, the tabu tenure is randomly selected within the range  $[t_0 + 1, t_{\max}]$ , following a uniform distribution. In the course of the implementation of intensification strategy, the tabu tenure is randomly selected within the range  $[t_{\min}, t_0]$ , following a uniform distribution. If there are no improvements in  $0.75I_{\text{imp}}$  iterations ( $I_{\text{imp}}$  means max number of consecutive nonimprovement iterations), the tabu tenure is randomly selected within the range  $[t_{\min}, t_{\max}]$ , following a uniform distribution.

**3.2.5. Neighborhood Trials Generation.** Trial solutions  $y^j$ ,  $j = 1, \dots, l$ , are generated by changing  $j$  positions in current solution  $x$  randomly based on tabu restrictions as in the following procedure.

*Procedure 1* ( $[y^1, \dots, y^l] = \text{Trials}(x, \text{TL}, l)$ ).

- (1) Repeat the following steps for  $j = 1, \dots, l$ .
- (2) Set  $y^j = x$ , and choose  $j$  random positions  $p_1, \dots, p_j$  of  $y^j$ .
- (3) Update the chosen positions by the rule  $y_{P_i}^j = 1 - y_{P_i}^j$ ,  $i = 1, \dots, j$ .
- (4) If  $y^j \in \text{TL}$ , then return to Step 2 to generate another  $y^j$ .

**3.2.6. Diversification Strategy.** The main roles of diversification strategy are to direct the search process to new solution regions and to accelerate escaping from local optima. ITSAR defines a vector  $v^F$  of dimension  $|C|$  which counts the numbers of choosing each condition attribute among the generated trial solutions. Then, a diverse solution  $x^{\text{div}}$  can be generated to contain attributes chosen with probability inversely proportional to their appearance in  $v^F$ . The procedure is as follows.

*Procedure 2* ( $[x^{\text{div}}] = \text{Diverse}(v^F)$ ).

- (1) Generate random numbers  $r_1, \dots, r_{|C|} \in (0, 1)$ .
- (2) Repeat the following step for  $i = 1, \dots, |C|$ .
- (3) If  $r_i > v_i^F / \sum_{i=1}^{|C|} v_i^F$ , set  $x^{\text{div}} = 1$ . Otherwise, set  $x^{\text{div}} = 0$ .

**3.2.7. Intensification Strategy.** If the search still cannot find any improvement during some iterations after generating  $x^{\text{div}}$ , ITSAR applies an intensification strategy to refine the best reduct  $x^{\text{best}}$  found so far. The best reduct  $x^{\text{best}}$  refinement, called Shaking, tries to reduce the attributes contained in  $x^{\text{best}}$  one by one without increasing  $\text{GK}(D \mid x^{\text{best}})$ . The search is continued from  $x^{\text{best}}$  no matter whether it can be improved through the Shaking Mechanism or not. Finally, the search process is terminated and a final refinement is applied. The procedure is as follows.

*Procedure 3* (Shaking( $x^{\text{best}}$ )).

- (1) Construct the set  $W$  of all positions of ones in  $x^{\text{best}}$ , that is, the elements of  $W$  represent the attributes contained in  $x^{\text{best}}$ .
- (2) Repeat the following steps for  $j = 1, \dots, |W|$ .
- (3) Delete the attribute  $w_j \in W$ , and compute a relative partition granularity.
- (4) Update  $x^{\text{best}}$  ( $x_{w_j}^{\text{best}} = 0$ ) if relative partition granularity is decreased or if relative partition granularity remains the same but the number of the attributes contained in reducts is decreased.

3.2.8. *ITSAR Algorithm*. The complete algorithm is as follows.

- (1) Let the Tabu List (TL) contain the two extreme solutions: solution of all ones and solution of all zeroes; set  $v^F$  to be a zero vector. Choose an initial solution  $x_0$ , and set the counter  $k = 0$ . Select  $I_{\text{max}}$ ,  $I_{\text{imp}}$ ,  $I_{\text{shak}}$ , and  $I_{\text{div}}$  such that  $I_{\text{max}} > I_{\text{imp}} > I_{\text{shak}} > I_{\text{div}}$ .
- (2) Generate neighborhood trials  $y^1, \dots, y^l$  around  $x^k$  using Procedure 1.
- (3) Set  $x^{k+1}$  equal to the best trial solution from  $y^1, \dots, y^l$ , and update TL,  $v^F$ , and  $x^{\text{best}}$ . Set  $k = k + 1$ .
- (4) If the number of iterations exceeds  $I_{\text{max}}$  or the number of iterations without improvement exceeds  $I_{\text{imp}}$ , terminate the search.
- (5) If the number of iterations without improvement exceeds  $I_{\text{shak}}$ , apply Procedure 3 to improve  $x^{\text{best}}$ , set  $x^k = x^{\text{best}}$ , and go to Step 2.
- (6) If the number of iterations without improvement exceeds  $I_{\text{div}}$ , apply Procedure 2 to obtain a new diverse solution  $x^{\text{div}}$ , set  $x^k = x^{\text{div}}$ , and go to Step 2.

3.3. *Value Reduction Based on Knowledge Granulation*. A heuristic algorithm based on knowledge granulation for value reduction, which is used to generate decision rules of group evacuation decision-making, is described as follows.

- (1) Examine the condition attribute of each decision rule by the column; if removing a condition attribute, three possible cases are as follows:
  - (1) if there are conflicting decision rules, then retain the dropped attribute value of conflicting decision rules, which means the value cannot be eliminated;
  - (2) if there are duplicate decision rules, then mark the dropped attribute value of duplicate decision rules as “\*”, which means the value can be eliminated;
  - (3) if there are no conflicting and duplicate decision rules, then mark the dropped attribute value as “?”, which means whether the value can be eliminated is pending.

- (2) Delete possible duplicate decision rules. If all the condition attributes of a decision rule are marked, then change the attribute value marked with “?” to the original attribute value.
- (3) Examine the attribute value marked with “?” of each decision rule.
  - (1) If there is only one “?”, go to (3); if there are more than one “?”, calculate the significance of all attribute values marked with “?” according to Definition 3.
  - (2) Select the attribute value marked with “?” and maximum of the significance in the decision rule
  - (3) If the decision can be made only by the attribute value without the mark, go to (4); otherwise, go to (5).
  - (4) Change the attribute value marked with “?” to “\*.”
  - (5) Change the attribute value marked with “?” to the original attribute value, and go to (2).
- (4) Delete decision rules in which all the condition attributes are marked as “\*” and the possible duplicate decision rules.
- (5) If there are two decision rules which satisfy the following two conditions: (a) only one condition attribute value is different, (b) one of different attribute values is marked as “\*”, then, for the decision rule in which different attribute values is marked as “\*”, if the decision can be made by the attribute values without the mark, delete another decision rule; otherwise, delete this rule.
- (6) Calculate the confidence of each rule; export the rules.

3.4. *Evaluation of the Approach*. In this study, examples were scarce; thus, cross-validation (CV) procedure [28] was used to evaluate the performance of the approach. A  $k$ -fold cross-validation procedure partitions the data into  $k$  disjoint subsets of nearly equal size. One of the subsets is reserved for testing, whereas the rest of the data constitute the training sample. This procedure is repeated  $k$  times. Each time using a different subset as the test set, and the final result is the arithmetic average of  $k$  separate tests.

We evaluated the performance of the approach by applying 10 times 5-fold cross-validation tests. The performance of the approach was measured by the hit rate of decision rule with maximum value of confidence ( $\text{hit}_0$ ), the hit rate of decision rule with maximum value of confidence and second largest value of confidence ( $\text{hit}_1$ ), and the comprehensive hit rate ( $\text{hit}_2$ ) computed by

$$\text{hit}_0 = \frac{c_0}{t}, \quad \text{hit}_1 = \frac{c_0 + c_1}{t}, \quad (8)$$

$$\text{hit}_2 = 0.6\text{hit}_0 + 0.4\text{hit}_1, \quad (9)$$

where  $c_0$  is the number of instances in the test set which can be correctly classified by decision rule with maximum value of

TABLE 3: An example of decision table with eight objects.

Objects	C1	C2	C3	C4	C5	C6	$D$
1	2	1	2	2	1	2	2
2	2	1	1	3	1	1	3
3	2	2	2	2	2	1	4
4	2	1	1	2	1	3	1
5	3	2	1	3	1	3	1
6	1	2	1	2	2	1	2
7	2	2	2	4	2	3	1
8	1	2	1	2	2	3	3

confidence,  $c_1$  is the number of instances in the test set which can be correctly classified by decision rule with second largest value of confidence,  $t$  is the total number of instances in the test set. The comprehensive hit rate ( $hit_2$ ) was also used as the model selection criterion.

#### 4. Application to Group Evacuation Decision-Making

This section presents our tests on group evacuation decision-making. We firstly develop the decision table in Section 4.1. Then we extract the reducts with the application of attribute reduction in Section 4.2. Based on the reducts, a set of decision rules are generated by means of value reduction in Section 4.3. Section 4.4 presents the testing results of the rules. Lastly, we compare the proposed method with other methods in performance.

*4.1. Representation of Decision Table.* The first step is to develop decision table for group evacuation decision-making. As discussed previously, we have used the dataset from evacuation behavior survey in Wuchang Railway Station area. The decision table includes 523 objects or samples. For each record, six conditional attributes are registered.

Table 3 shows eight objects of group evacuation decision-making used as the example of decision table.

*4.2. Reduction of Attributes.* The algorithms for attribute reduction and value reduction were programmed in MATLAB and applied to the decision table of group evacuation decision-making. The parameter values used in ITSAR were set to the following values:  $t = [5, 10]$ ,  $l = 3$ ,  $I_{\max} = 100$ ,  $I_{\text{imp}} = 40$ ,  $I_{\text{shak}} = 20$ ,  $I_{\text{div}} = 10$ . These chosen values are based on the common setting in [17, 29]. This indicates the potential for future improvement of ITSAR by systematically fine-tuning these parameters using statistical tests as suggested by Xu et al. [30].

After attribute reduction by applying 10 times 5-fold cross-validation tests, some reducts can be obtained. Table 4 shows the frequency of individual reduct occurring among the set of reducts. Evidently, there is one common attribute, that is, group relationship ( $C6$ ), occurring in all the reducts.

TABLE 4: Reducts and their frequency.

Reducts	Frequency
{C1, C4, C6}	8
{C2, C3, C6}	2
{C2, C4, C6}	14
{C4, C6}	11
{C2, C3, C4, C6}	6
{C1, C2, C4, C6}	5
{C3, C4, C6}	3
{C1, C6}	1

*4.3. Decision Rules.* Based on reducts obtained in the previous step, decision rules can be generated from the decision table by value reduction. For the reduct with the highest frequency, rules are obtained from the corresponding training set and shown in Table 5. In these decision rules, 2 rules are certain and others are uncertain. For example, rule 1 and rule 3 are selected to describe below.

Rule 1 means that if condition attribute values satisfy the following conditions, that is, gender is male and temperament is choleric, and group relationship is lover, then group decision-making mode chosen by individual is choosing the route approved by the one doing things reasonably. The confidence of this rule is 1.

Rule 3 means that if condition attribute values satisfy the following conditions, that is, gender is female and temperament is choleric, and group relationship is lover, then group decision-making mode chosen by individual has three possibilities, that is, choosing the route approved by self (with the confidence of 0.125), or choosing the route approved by the one familiar with the route (with the confidence of 0.5), or choosing the route approved by the one doing things reasonably (with the confidence of 0.375).

*4.4. Results of Testing.* Decision rules generated from the training set are applied to the corresponding testing set in order to harvest a performance estimate. The results from 10 times 5-fold cross-validation tests show that the range of  $hit_0$  is from 0.298 to 0.509 and the range of  $hit_1$  is from 0.529 to 0.759. As shown in Figure 1, the majority of  $hit_0$  is above 0.4 and the majority of  $hit_1$  is above 0.6. For

TABLE 5: Decision rules of the corresponding sample.

Decision rules	Confidence
Rule 1: If (C2 = 2) and (C4 = 1) and (C6 = 2), then (D = 4)	1
Rule 2: If (C2 = 2) and (C4 = 4) and (C6 = 2), then (D = 3)	1
Rule 3: If (C2 = 1) and (C4 = 1) and (C6 = 2), then (D = 2 or 3 or 4)	(0.125, 0.5, 0.375)
Rule 4: If (C4 = 2), then (D = 1 or 2 or 3 or 4)	(0.256, 0.185, 0.385, 0.174)
Rule 5: If (C2 = 1) and (C6 = 1), then (D = 1 or 2 or 3 or 4)	(0.241, 0.177, 0.418, 0.164)
Rule 6: If (C2 = 1) and (C4 = 3) and (C6 = 2), then (D = 1 or 2 or 4)	(0.286, 0.428, 0.286)
Rule 7: If (C2 = 1) and (C4 = 4) and (C6 = 2), then (D = 2 or 4)	(0.333, 0.667)
Rule 8: If (C2 = 1) and (C4 = 4) and (C6 = 3), then (D = 1 or 2 or 4)	(0.25, 0.5, 0.25)
Rule 9: If (C2 = 2) and (C4 = 3), then (D = 1 or 2 or 3 or 4)	(0.13, 0.208, 0.532, 0.13)
Rule 10: If (C2 = 2) and (C4 = 4) and (C6 = 1), then (D = 1 or 3 or 4)	(0.143, 0.571, 0.286)
Rule 11: If (C2 = 2) and (C4 = 4) and (C6 = 3), then (D = 1 or 2 or 4)	(0.167, 0.167, 0.666)
Rule 12: If (C4 = 1) and (C6 = 1), then (D = 1 or 2 or 3 or 4)	(0.148, 0.185, 0.519, 0.148)
Rule 13: If (C4 = 1) and (C6 = 3), then (D = 1 or 2 or 3 or 4)	(0.207, 0.276, 0.31, 0.207)
Rule 14: If (C4 = 3) and (C6 = 3), then (D = 1 or 2 or 3 or 4)	(0.175, 0.222, 0.539, 0.064)

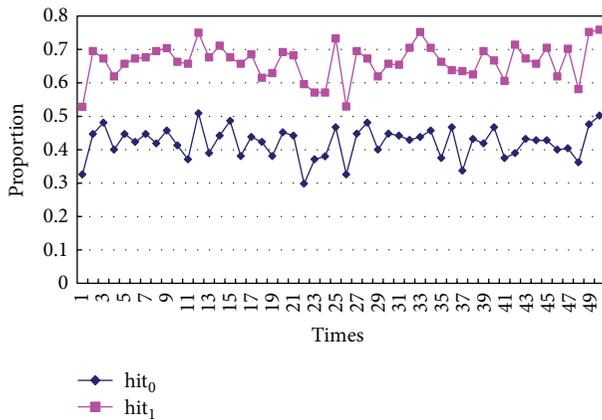


FIGURE 1: Performance of rough set theory.

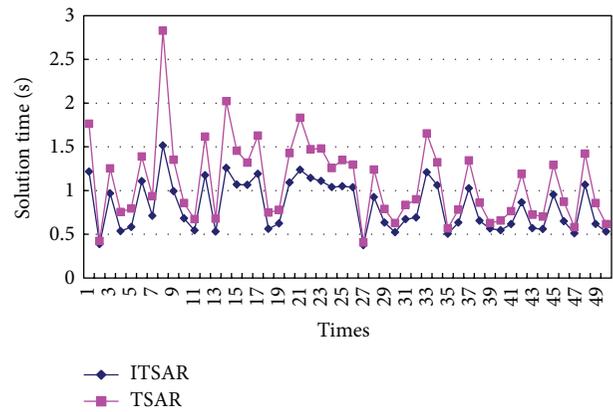


FIGURE 2: Solution times of ITSAR and TSAR.

the best rough set model,  $hit_0$ ,  $hit_1$ , and  $hit_2$  are 0.502, 0.759, and 0.605, respectively. The best rough set model consists of three conditional attributes: gender, temperament, and group relationship.

4.5. *Comparison with Other Methods.* To get a better picture of the power of rough set theory, a comparison with other techniques using the same training and testing samples would prove useful. For the purpose of comparison, we chose tabu search for attribute reduction (TSAR) [17] and multinomial logistics regression (MLR) [31].

In the TSAR algorithm, the dependency degree of decision attribute is used to measure the quality of a solution, and fixed tabu tenure is used. We set fixed tabu tenure as 8 in our study. TSAR and ITSAR could obtain the same reducts in this paper. The solution times of two methods for 50 runs are displayed in Figure 2. Regarding average solution time, ITSAR obtained the reducts in 24.7% less time than TSAR. The average solution times were 0.827 CPU seconds for ITSAR and 1.1 CPU seconds for TSAR.

Multinomial logistics regression can be used when a categorical dependent variable has more than two categories. For the implementation of the multinomial logistic regression model, the backward elimination procedure was performed by using SPSS software in this study. The performance of multinomial logistics regression model was determined by cross-validation procedure described in Section 3.4. Here, in (8),  $c_0$  is the number of instances in the test set which can be correctly classified by the category with maximum value of probability, and  $c_1$  is the number of instances in the test set which can be correctly classified by the category with the second largest value of probability. The best logistic regression model was extracted on the basis of goodness-of-fit test and the hit rates.

The results from 10 times 5-fold cross-validation test show that the range of  $hit_0$  is from 0.123 to 0.48 and the range of  $hit_1$  is from 0.451 to 0.723. As shown in Figure 3, the majority of  $hit_0$  is below 0.4 and the majority of  $hit_1$  is above 0.6. For the best logistic regression model,  $hit_0$ ,  $hit_1$ , and  $hit_2$  are 0.48, 0.693, and 0.565, respectively. Goodness-of-fit measures of the best logistic regression model are displayed in Table 6.

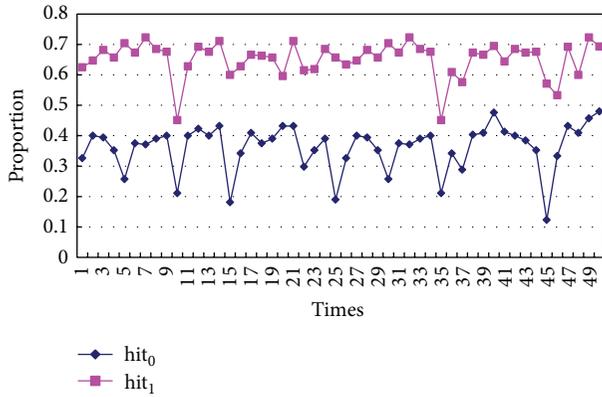


FIGURE 3: Performance of multinomial logistics regression.

TABLE 6: Goodness-of-fit measures of the best logistic regression model.

	$\chi^2$	DF	Sig.
Pearson	401.191	360	0.066
Deviance	385.701	360	0.168

TABLE 7: Performance comparison.

	hit <sub>0</sub>	hit <sub>1</sub>
Rough set theory	0.421	0.663
MLR	0.362	0.651

The best logistic regression model consists of three variables: gender, temperament, and group relationship, which is the same as conditional attributes in the best rough set model.

Table 7 depicts the average results of 10 times 5-fold cross-validation test for two approaches. The results of the approach comparison show that rough set theory method is superior to multinomial logistics regression in terms of testing performance.

On the other hand, the fluctuation of the curve in Figure 1 is relatively moderate, whereas the fluctuation of the curve in Figure 3 is relatively marked. The comparisons suggest that the stability of testing performance for rough set theory method is better than that for multinomial logistics regression.

## 5. Conclusions

In this paper, we focus on the analysis of group evacuation decision-making in the face of disagreement among decision makers in passenger transport hub area. Rough set theory is applied to analyze group evacuation decision-making. Based on evacuation behavior survey, we develop the decision table of group evacuation decision-making. An improved tabu search-based attribute reduction (ITSAR) is proposed to find a minimal reduct of decision table, and then a heuristic algorithm based on knowledge granulation for value reduction is introduced for rule extraction of decision table. According to the presented research, rules of

group evacuation decision-making are generated in a readily understandable form (a set of simple if-then statements). By using 10 times 5-fold cross-validation tests, we compare the proposed method with other methods in performance. The results show that ITSAR outperformed TSAR in terms of solution time, and rough set theory has the advantage over multinomial logistics regression for the analysis of group evacuation decision-making. It can be concluded that rough set theory can quickly obtain more simple decision rules of group evacuation decision-making and provide a new way for evacuation behavior analysis.

Further research mainly includes two aspects. First, it is worthwhile to develop effective update method for rule database after increasing new samples. Second, this model could be integrated with a larger set of behavioral models into an agent-based simulation framework to comprehensively model the evacuation process, which would help public agencies develop evacuation plans that align with evacuee choices and behavior.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

Research for this study was supported by the National Natural Science Foundation of China (51208400, 51108362, and 51078299) and the Fundamental Research Funds for the Central Universities of China (no. 2011-IV-125).

## References

- [1] M. K. Lindell, J.-C. Lu, and C. S. Prater, "Household decision making and evacuation in response to Hurricane Lili," *Natural Hazards Review*, vol. 6, no. 4, pp. 171–179, 2005.
- [2] N. Dash and H. Gladwin, "Evacuation decision making and behavioral responses: individual and household," *Natural Hazards Review*, vol. 8, no. 3, pp. 69–77, 2007.
- [3] J. Czajkowski, "Is it time to go yet? Understanding household hurricane evacuation decisions from a dynamic perspective," *Natural Hazards Review*, vol. 12, no. 2, pp. 72–84, 2011.
- [4] S. Hasan, S. Ukkusuri, H. Gladwin, and P. Murray-Tuite, "Behavioral model to understand household-level hurricane evacuation decision making," *Journal of Transportation Engineering*, vol. 137, no. 5, pp. 341–348, 2011.
- [5] M. K. Lindell, J. E. Kang, and C. S. Prater, "The logistics of household hurricane evacuation," *Natural Hazards*, vol. 58, no. 3, pp. 1093–1109, 2011.
- [6] R. Mesa-Arango, S. Hasan, S. V. Ukkusuri, and P. Murray-Tuite, "Household-Level model for hurricane evacuation destination type choice using hurricane ivan data," *Natural Hazards Review*, vol. 14, no. 1, pp. 11–20, 2013.
- [7] S. M. Lo, M. Liu, P. H. Zhang, and R. K. K. Yuen, "An artificial neural-network based predictive model for pre-evacuation human response in domestic building fire," *Fire Technology*, vol. 45, no. 4, pp. 431–449, 2009.

- [8] C. M. Zhao, S. M. Lo, S. P. Zhang, and M. Liu, "A post-fire survey on the pre-evacuation human behavior," *Fire Technology*, vol. 45, no. 1, pp. 71–95, 2009.
- [9] H. Fu and C. G. Wilmot, "Sequential logit dynamic travel demand model for hurricane evacuation," *Transportation Research Record*, no. 1882, pp. 19–26, 2004.
- [10] Y. Hsu and S. Peeta, "An aggregate approach to model evacuee behavior for no-notice evacuation operations," *Transportation*, vol. 40, no. 3, pp. 671–696, 2013.
- [11] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.
- [12] P. Sawicki and J. Zak, "Technical diagnostic of a fleet of vehicles using rough set theory," *European Journal of Operational Research*, vol. 193, no. 3, pp. 891–903, 2009.
- [13] J.-Y. Shyng, H.-M. Shieh, G.-H. Tzeng, and S.-H. Hsieh, "Using FSBT technique with Rough Set Theory for personal investment portfolio analysis," *European Journal of Operational Research*, vol. 201, no. 2, pp. 601–607, 2010.
- [14] L.-Y. Zhai, L.-P. Khoo, and S.-C. Fok, "Feature extraction using rough set theory and genetic algorithms—an application for the simplification of product quality evaluation," *Computers and Industrial Engineering*, vol. 43, no. 4, pp. 661–676, 2002.
- [15] R. Jensen and Q. Shen, "Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 12, pp. 1457–1471, 2004.
- [16] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 459–471, 2007.
- [17] A.-R. Hedar, J. Wang, and M. Fukushima, "Tabu search for attribute reduction in rough set theory," *Soft Computing*, vol. 12, no. 9, pp. 909–918, 2008.
- [18] A.-E. Hassanien, "Rough set approach for attribute reduction and rule generation: a case of patients with suspected breast cancer," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 11, pp. 954–962, 2004.
- [19] X. Bai, M. Zhang, Y. Qiu, and Q. Wu, "Algorithm for decision rules reduction in incomplete information system based on binary discernibility matrix," in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA '09)*, pp. 4061–4066, August 2009.
- [20] L. Zhang, X. Lu, H. Wu, and S. Hao, "Improved heuristic algorithm used in attribute value reduction of rough set," *Chinese Journal of Scientific Instrument*, vol. 30, no. 1, pp. 82–85, 2009.
- [21] J. Liang, J. Wang, and Y. Qian, "A new measure of uncertainty based on knowledge granulation for rough sets," *Information Sciences*, vol. 179, no. 4, pp. 458–470, 2009.
- [22] E. Rolland, D. A. Schilling, and J. R. Current, "An efficient tabu search procedure for the p-Median Problem," *European Journal of Operational Research*, vol. 96, no. 2, pp. 329–342, 1997.
- [23] P. Yin and Y. Guo, "Optimization of multi-criteria website structure based on enhanced tabu search and web usage mining," *Applied Mathematics and Computation*, vol. 219, no. 24, pp. 11082–11095, 2013.
- [24] Q. Wu and J.-K. Hao, "An adaptive multistart tabu search approach to solve the maximum clique problem," *Journal of Combinatorial Optimization*, vol. 26, no. 1, pp. 86–108, 2013.
- [25] Q. Feng, D. Miao, and Y. Cheng, "An approach to knowledge reduction based on relative partition granularity," in *Proceedings of the IEEE International Conference on Granular Computing (GRC '08)*, pp. 226–231, August 2008.
- [26] F. Glover, "Future paths for integer programming and links to artificial intelligence," *Computers & Operations Research*, vol. 13, no. 5, pp. 533–549, 1986.
- [27] C. Xie and M. A. Turnquist, "Lane-based evacuation network optimization: an integrated Lagrangian relaxation and tabu search approach," *Transportation Research C: Emerging Technologies*, vol. 19, no. 1, pp. 40–63, 2011.
- [28] P. A. Berger, "Rough set rule induction for suitability assessment," *Environmental Management*, vol. 34, no. 4, pp. 546–558, 2004.
- [29] C. K. Y. Lin and R. C. W. Kwok, "Multi-objective metaheuristics for a location-routing problem with multiple use of vehicles on real data and simulated data," *European Journal of Operational Research*, vol. 175, no. 3, pp. 1833–1849, 2006.
- [30] J. Xu, S. Y. Chiu, and F. Glover, "Fine-tuning a tabu search algorithm with statistical tests," *International Transactions in Operational Research*, vol. 5, no. 3, pp. 233–244, 1998.
- [31] G. H. Bham, B. S. Javvadi, and U. R. R. Manepalli, "Multinomial logistic regression model for Single-Vehicle and multivehicle collisions on urban u.s. Highways in arkansas," *Journal of Transportation Engineering-Asce*, vol. 138, no. 6, pp. 786–797, 2012.

## Research Article

# A Bayesian Network Model for Origin-Destination Matrices Estimation Using Prior and Some Observed Link Flows

**Lin Cheng, Senlai Zhu, Zhaoming Chu, and Jingxu Cheng**

*School of Transportation, Southeast University, Nanjing 210096, China*

Correspondence should be addressed to Senlai Zhu; zhusenlai@163.com

Received 1 November 2013; Accepted 11 March 2014; Published 13 April 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Lin Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a Bayesian network model for estimating origin-destination matrices. Most existing Bayesian methods adopt prior OD matrixes, which are always troublesome to be obtained. Since transportation systems normally have stored large amounts of historical link flows, a Bayesian network model using these prior link flows is proposed. Based on some observed link flows, the estimation results are updated. Under normal distribution assumption, the proposed Bayesian network model considers the level of total traffic flow, the variability of link flows, and the violation of the traffic flow conservation law. Both the point estimation and the corresponding probability intervals can be provided by this model. To solve the Bayesian network model, a specific procedure which can avoid matrix inversion is proposed. Finally, a numerical example is given to illustrate the proposed Bayesian network method. The results show that the proposed method has a high accuracy and practical applicability.

## 1. Introduction

Information about the traffic demand, which commonly means the origin-destination (OD) matrices, has been traditionally used by transportation planning agencies to evaluate the impact of various strategic transportation plans. And the real-time OD matrices are essential for real-time traffic applications, especially in the intelligent transportation system (ITS), such as the real-time route guidance via a dynamic traffic assignment or the evaluation of various ITS deployment alternatives [1, 2].

Various methods have been proposed to estimate OD matrices by using aggregate data such as OD demand counts and/or a set of traffic counts observed on the links. Using information derived from traffic counts is very attractive because they are cheap, easy, and immediate data. However, based on these data, we cannot obtain a unique OD matrix because the number of OD pairs is much larger than the number of links in large-scale transportation networks and there are infinite solutions satisfying the conservation law.

In order to have a unique solution which must be close to the actual one, one has to give more information. Normally, people use a prior OD matrix which can be obtained by many different methods, such as an old out-of-date or subjectively

guessed OD matrix. These methods for estimating OD matrices can be classified as (1) least squares [3–9] and generalized least squares [10–12] methods, (2) entropy or information based methods [13, 14], and (3) statistical based methods.

Providing variability information of the traffic flow estimation is the most important advantage of the statistical methods. Normally, other methods give only the particular values of the OD and link flows, while statistical methods could also provide the corresponding probability intervals. The statistical methods can be categorized as follows. (1) Classical methods [15–17]: the traffic flows are assumed multivariate random variables given some parametric families, such as Poisson, Gamma, and multivariate normal. Then, the problem reduces to estimating the parameters and becomes a standard statistical problem. (2) Bayesian methods [18–21]: these methods also consider parametric families of distributions, but the parameters are considered as random variables themselves. Particularly, among the Bayesian methods, using Bayesian network [22–24] can easily know the relationship of all the variables (link flows and OD flows) and then simplify the calculation.

Whether using prior information (historical information or experience) or not is the main difference between Bayesian methods and classical statistical methods. In the Bayesian

methods, based on some prior information, the prior distribution of some parameters or variables can be determined. Then by updating the sample information (observed information), we can derive the posterior distribution, which is the fundamental inferential tool of the Bayesian methods.

Generally, the quality of the prior information can affect the accuracy of the estimation when using the Bayesian methods. The prior information used by almost all existing Bayesian methods for estimating OD matrices is a prior OD matrix. However, it is difficult to guarantee the accuracy of the prior OD matrix, which is outdated or subjectively guessed. Moreover, it is even impossible to get a prior OD matrix in some cases, especially in a newly developed city.

In reality, there are usually large amounts of historical link flow data stored in the cities' transportation system data base. Compared with a prior OD matrix, prior (historical) link flows are more accurate as they were obtained by traffic detectors or manual investigation. Therefore, in this paper, in order to estimate OD matrices, we propose a Bayesian network (BN) method using prior link flows and a set of new observed link flows. Based on these prior link flows, we can derive the prior distribution of link flows and OD flows. Then, by updating a set of observed link flows, we can modify the means and reduce the variances of the remaining variables. Using these updated means and variances, we can obtain the posterior distribution of all the variables. Based on the posterior distribution, both the point estimation and the corresponding probability intervals can be provided.

Note that the level of total traffic flow varies randomly and deterministically in similar situations (vacation, peak hour, special weather conditions, etc.) [20, 24]. So the proposed BN model also considers the level of total traffic flow, which is very useful for many real-time traffic applications. In addition, the BN model also considers the variability of link flows and the violation of the conservation law.

The rest of the paper is organized as follows. Section 2 briefly introduces Bayesian network and Gaussian Bayesian networks. In Section 3, the proposed BN model for estimating OD matrices and its main assumptions are described. In Section 4, using the Bayesian network model, a specific procedure for estimating OD matrices is proposed. In Section 5, a numerical example is provided to illustrate the proposed model and clarify some of its implementation details. Finally, some conclusions are provided in Section 6.

## 2. Bayesian Network and Gaussian Bayesian Network

In this section, we briefly review the Bayesian network and Gaussian Bayesian network, which are the basic tools of this paper.

*Definition 1* (Bayesian network). A Bayesian network is a pair  $(\mathbf{G}, \mathbf{P})$ , where  $\mathbf{G}$  is a directed acyclic graph (DAG) defined on a set of nodes  $\mathbf{X}$ ,  $\mathbf{P} = \{p(x_1 | \boldsymbol{\pi}_1), \dots, p(x_n | \boldsymbol{\pi}_n)\}$  is a set of  $n$  conditional probability densities (CPDs), and  $\boldsymbol{\pi}_i$  is the set of

parents of node  $X_i$  in  $\mathbf{G}$ . The set  $\mathbf{P}$  defines the associated joint probability density (JPD) as

$$p(\mathbf{X}) = \prod_{i=1}^n p(x_i | \boldsymbol{\pi}_i). \quad (1)$$

The graph  $\mathbf{G}$  contains all the qualitative information about the relationships among the variables. As a supplement, the probabilities in  $\mathbf{P}$  quantify the qualitative information in graph  $\mathbf{G}$ .

In Bayesian networks, the factorization of JPD implied by (1) is normally very simple and the conditional independence relations among variables can be inferred directly from the graph  $\mathbf{G}$ , which makes the evidence propagation easy. Due to these advantages, Bayesian network models have been used widely to solve a large variety of practical problems [25, 26].

Bayesian networks can be applied to many distributions. For the sake of illustration, we consider the important and particular case of Gaussian Bayesian networks, in which the traffic flows distribution is supposed to be a normal distribution. A normal distribution for traffic flows is reasonable, because these random variables are the sum of a great number of independent Bernoulli experiments in which the users decide where to travel and which routes to choose. In the literature, Gaussian Bayesian networks have been used frequently [24, 27].

*Definition 2* (Gaussian Bayesian network). A Bayesian network  $(\mathbf{G}, \mathbf{P})$  is said to be a Gaussian Bayesian network if and only if the joint probability distribution (JPD) associated with its variables  $\mathbf{X}$  is a multivariate normal distribution,  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , that is, with joint probability density function:

$$f(\mathbf{X}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right\}, \quad (2)$$

where  $\boldsymbol{\mu}$  is the mean vector,  $\boldsymbol{\Sigma}$  is the  $n \times n$  covariance matrix,  $|\boldsymbol{\Sigma}|$  is the determinant of  $\boldsymbol{\Sigma}$ , and  $\boldsymbol{\mu}^T$  is the transpose of  $\boldsymbol{\mu}$ .

The JPD of the variables in a Gaussian Bayesian network can be specified as in (1) by the product of a set of CPDs, whose joint probability density function is

$$f(x_i | \boldsymbol{\pi}_i) \sim N \left( \mu_i + \sum_{j=1}^{i-1} \beta_{ij} (x_j - \mu_j), \psi_i^2 \right), \quad (3)$$

where  $\beta_{ij}$  is the regression coefficient of  $X_j$  in the regression of  $X_i$  on its parents  $\boldsymbol{\pi}_i$ .

And the conditional variance of  $X_j$  is

$$\psi_i^2 = \Sigma_i - \boldsymbol{\Sigma}_{i\boldsymbol{\pi}_i} \boldsymbol{\Sigma}_{\boldsymbol{\pi}_i}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\pi}_i}^T, \quad (4)$$

where  $\Sigma_i$  is the unconditional variance of  $X_i$ ,  $\boldsymbol{\Sigma}_{i\boldsymbol{\pi}_i}$  is the covariance matrix between  $X_i$  and the variables in  $\boldsymbol{\pi}_i$ , and  $\boldsymbol{\Sigma}_{\boldsymbol{\pi}_i}$  is the covariance matrix of  $\boldsymbol{\pi}_i$ .

## 3. Proposed Bayesian Network Model

Since Bayesian network has so many advantages as introduced in Section 2, in this section we propose a Bayesian

network (BN) model to reproduce the probabilistic structure of link and OD flows.

**3.1. Model Assumptions.** Assuming we have some prior (historical) link flows, in order to give the prior distribution of the link flows, we make the following assumptions.

*Assumption 3.* The link flows are given by

$$\mathbf{V} = \mathbf{K}\mathbf{U} + \boldsymbol{\eta}. \quad (5)$$

*Assumption 4.* The variable  $U$  is a normal random variable with mean  $\mu_U$  and variance-covariance matrix  $\sigma_U^2$ , where  $U$  is a normal random variable and measures the level of total mean flow. It reflects that traffic flows vary randomly and deterministically in similar situations (vacation, peak hour, special weather conditions, etc.).  $\mathbf{K}$  is a vector, whose elements measure the relative weights of link flows with respect to the total traffic flow;  $\boldsymbol{\eta}$  is a vector of independent normal random variables with zero mean; and  $\eta_a$  measures the discrepancy of the flow of link  $a$  with respect to its mean.

Note that traffic flows vary randomly and deterministically in similar situations (vacation, peak hour, special weather conditions, etc.) [20, 24]. Assumption 3 can take this into account. The distribution of  $U$  varies with the situation. Based on prior link flows and considering the similar situation, the distribution of  $U$  and the initial vector  $\mathbf{K}$  are determined. Then we can easily derive the prior distribution of link flows, which will be shown later. Assumption 4 is a normal assumption, which is also adopted in Maher [18], Hazelton [20], Castillo et al. [24], and so forth.

To give the prior distribution of OD flows, we first consider the well-known conservation law equation:

$$V_a = \sum_i \left( \sum_k T_i p_{ik} \delta_{ak}^i \right) = \sum_i T_i \left( \sum_k p_{ik} \delta_{ak}^i \right), \quad (6)$$

where  $T_i$  and  $V_a$  are the flows of OD pair  $i$  and link  $a$ , respectively.  $\delta_{ak}^i$  is the incidence element; that is, it takes value 1 if link  $a$  belongs to route  $k$  of OD pair  $i$  and 0, otherwise.  $p_{ik}$  is the proportion of users from OD pair  $i$  choosing route  $k$ . In this paper, the route choice proportions are defined by a logit model as follows:

$$p_{ik} = \frac{\exp(-\theta c_k^i)}{\sum_{l \in R_i} \exp(-\theta c_l^i)}, \quad (7)$$

where  $\theta$  is a parameter measuring travelers' sensitivity to the cost difference between routes;  $c_k^i$  is the cost associated with route  $k$  of OD pair  $i$ .

Equation (6) can be written as

$$V_a = \sum_i \left( \sum_k p_{ik} \delta_{ak}^i \right) T_i. \quad (8)$$

Set  $d_{ai} = \sum_k p_{ik} \delta_{ak}^i$  to represent the proportions of users from OD pair  $i$  choosing link  $a$ . Then (8) can be rewritten in the form of matrix as

$$\mathbf{V} = \mathbf{D}\mathbf{T}. \quad (9)$$

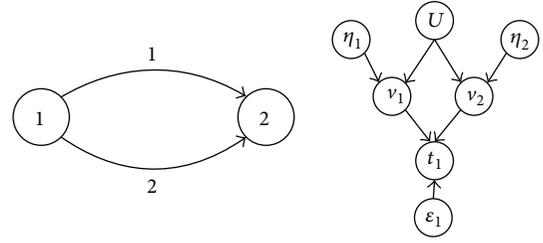


FIGURE 1: Bayesian network model.

Matrix  $\mathbf{D}$  is not necessarily reversible because it is not necessarily a square. So we do the following conversion of (9):

$$\mathbf{D}^T \mathbf{V} = \mathbf{D}^T \mathbf{D} \mathbf{T}. \quad (10)$$

If matrix  $\mathbf{D}^T \mathbf{D}$  is of full rank, it is reversible. Then, (10) can be written as

$$\mathbf{T} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}. \quad (11)$$

Set  $\boldsymbol{\beta} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T$  and according to (11), we make the following assumption.

*Assumption 5.* The OD flows are given by

$$\mathbf{T} = \boldsymbol{\beta} \mathbf{V} + \boldsymbol{\varepsilon}, \quad (12)$$

where  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$  are mutually independent normal random variables with mean  $E(\varepsilon_i)$  and variance  $\psi_i^2$ . The variables  $\boldsymbol{\varepsilon}$  represent OD flows apart from those using links of the considered network. Setting all the variables of  $\boldsymbol{\varepsilon}$  to be null or evaluating their values, the conservation law equation can be satisfied.

**3.2. The Complete Model.** Based on Definitions 1 and 2, in order to complete our BN model, we need to define an associated graph  $\mathbf{G}$ . For example, consider the simple network shown in the left of Figure 1, which has 2 nodes, 2 links, and 1 OD pair 1-2. The right of Figure 1 shows the associated Bayesian network. The link flow node  $v_a$  has as parents the corresponding node  $U$  and  $\eta_a$ . The OD flow node  $t_i$  has as parents the corresponding node  $v_a$  and  $\varepsilon_i$ .

Then according to Assumptions 3 and 4, we get the variance-covariance matrix of  $\mathbf{V}$ :

$$\boldsymbol{\Sigma}_{\mathbf{V}} = (\mathbf{K} \ \mathbf{I}) \boldsymbol{\Sigma}_{(U, \boldsymbol{\eta})} \begin{pmatrix} \mathbf{K}^T \\ \mathbf{I} \end{pmatrix} = \sigma_U^2 \mathbf{K} \mathbf{K}^T + \mathbf{D}_{\boldsymbol{\eta}}, \quad (13)$$

where  $\boldsymbol{\Sigma}_{(U, \boldsymbol{\eta})}$  and  $\mathbf{D}_{\boldsymbol{\eta}}$  are diagonal matrices.  $\mathbf{D}_{\boldsymbol{\eta}}$  is the variance-covariance matrix of  $\boldsymbol{\eta}$ .

Based on Assumption 5, we get

$$\begin{pmatrix} \mathbf{V} \\ \mathbf{T} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{V} \\ \boldsymbol{\varepsilon} \end{pmatrix}. \quad (14)$$

Then, the mean  $E[(\mathbf{V}, \mathbf{T})]$  is

$$E[(\mathbf{V}, \mathbf{T})] = \begin{pmatrix} E(U) \mathbf{K} \\ E(U) \boldsymbol{\beta} \mathbf{K} + E(\boldsymbol{\varepsilon}) \end{pmatrix}. \quad (15)$$

And the variance-covariance matrix of  $(\mathbf{V}, \mathbf{T})$  is

$$\Sigma_{(\mathbf{V}, \mathbf{T})} = \begin{pmatrix} \Sigma_{\mathbf{V}} & \Sigma_{\mathbf{V}} \boldsymbol{\beta}^T \\ \boldsymbol{\beta} \Sigma_{\mathbf{V}} & \boldsymbol{\beta} \Sigma_{\mathbf{V}} \boldsymbol{\beta}^T + \mathbf{D}_{\boldsymbol{\varepsilon}} \end{pmatrix}, \quad (16)$$

where  $\mathbf{D}_{\boldsymbol{\varepsilon}}$  is the variance matrix of  $\boldsymbol{\varepsilon}$ .

In summary, all random variables involved in our model are related by the linear expression:

$$\begin{pmatrix} U \\ \boldsymbol{\eta} \\ \boldsymbol{\varepsilon} \\ \mathbf{V} \\ \mathbf{T} \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{K} & \mathbf{I} & \mathbf{0} \\ \boldsymbol{\beta} \mathbf{K} & \boldsymbol{\beta} & \mathbf{I} \end{pmatrix} \begin{pmatrix} U \\ \boldsymbol{\eta} \\ \boldsymbol{\varepsilon} \end{pmatrix}. \quad (17)$$

The mean  $E[(U, \boldsymbol{\eta}, \boldsymbol{\varepsilon}, \mathbf{V}, \mathbf{T})]$  is

$$\begin{pmatrix} E(U) \\ E(\boldsymbol{\eta}) \\ E(\boldsymbol{\varepsilon}) \\ E(\mathbf{V}) \\ E(\mathbf{T}) \end{pmatrix} = \begin{pmatrix} E(U) \\ \mathbf{0} \\ E(\boldsymbol{\varepsilon}) \\ E(U) \mathbf{K} \\ E(U) \boldsymbol{\beta} \mathbf{K} + E(\boldsymbol{\varepsilon}) \end{pmatrix}. \quad (18)$$

And the variance-covariance matrix  $\Sigma_{(U, \boldsymbol{\eta}, \boldsymbol{\varepsilon}, \mathbf{V}, \mathbf{T})}$  is

$$\Sigma = \begin{pmatrix} \sigma_U^2 & \mathbf{0} & \mathbf{0} & \sigma_U^2 \mathbf{K}^T & \sigma_U^2 \mathbf{K}^T \boldsymbol{\beta}^T \\ \mathbf{0} & \mathbf{D}_{\boldsymbol{\eta}} & \mathbf{0} & \mathbf{D}_{\boldsymbol{\eta}} & \mathbf{D}_{\boldsymbol{\eta}} \boldsymbol{\beta}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{\boldsymbol{\varepsilon}} & \mathbf{0} & \mathbf{D}_{\boldsymbol{\varepsilon}} \\ \sigma_U^2 \mathbf{K} & \mathbf{D}_{\boldsymbol{\eta}} & \mathbf{0} & \sigma_U^2 \mathbf{K} \mathbf{K}^T + \mathbf{D}_{\boldsymbol{\eta}} & (\sigma_U^2 \mathbf{K} \mathbf{K}^T + \mathbf{D}_{\boldsymbol{\eta}}) \boldsymbol{\beta}^T \\ \sigma_U^2 \boldsymbol{\beta} \mathbf{K} & \boldsymbol{\beta} \mathbf{D}_{\boldsymbol{\eta}} & \mathbf{D}_{\boldsymbol{\varepsilon}} & \boldsymbol{\beta} (\sigma_U^2 \mathbf{K} \mathbf{K}^T + \mathbf{D}_{\boldsymbol{\eta}}) & (\sigma_U^2 \mathbf{K} \mathbf{K}^T + \mathbf{D}_{\boldsymbol{\eta}}) \boldsymbol{\beta}^T + \mathbf{D}_{\boldsymbol{\varepsilon}} \end{pmatrix}. \quad (19)$$

Then, the prior distribution (joint probability density function) of all the variables can be given as

$$\begin{aligned} & f(V_1, V_2, \dots, V_m, T_1, T_2, \dots, T_n) \\ &= f_{N(\boldsymbol{\mu}_{\mathbf{V}}, \Sigma_{\mathbf{V}})}(V_1, V_2, \dots, V_m) \prod_{i=1}^n p\left(T_i \mid \prod_i = \boldsymbol{\pi}_i\right). \end{aligned} \quad (20)$$

#### 4. Estimating OD Matrices Using the Proposed BN Model

In this section, using the proposed BN model, we describe how to estimate the OD matrices when some new observed link flows are available.

Since we have obtained the prior distribution of all the variables, we can use the following equations to update the mean and the covariance matrix of the variables [23, 24] based on some observed variables. Note that one only needs to consider the unobserved variables conditioned on the

observed variables and then update the expected values and covariance of the remaining variables. These equations are

$$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{Z}=z} = \boldsymbol{\mu}_{\mathbf{Y}} + \Sigma_{\mathbf{YZ}} \Sigma_{\mathbf{ZZ}}^{-1} (z - \boldsymbol{\mu}_{\mathbf{Z}}), \quad (21)$$

$$\Sigma_{\mathbf{Y}|\mathbf{Z}=z} = \Sigma_{\mathbf{YY}} - \Sigma_{\mathbf{YZ}} \Sigma_{\mathbf{ZZ}}^{-1} \Sigma_{\mathbf{ZY}}, \quad (22)$$

$$\boldsymbol{\mu}_{\mathbf{Z}|\mathbf{Z}=z} = z, \quad (23)$$

$$\Sigma_{\mathbf{Z}|\mathbf{Z}=z} = 0, \quad (24)$$

where  $\mathbf{Y}$  and  $\mathbf{Z}$  are the sets of unobserved and observed variables, respectively;  $\boldsymbol{\mu}_{\mathbf{Y}}$  and  $\Sigma_{\mathbf{YY}}$  are the mean vector and covariance matrix of  $\mathbf{Y}$ ;  $\boldsymbol{\mu}_{\mathbf{Z}}$  and  $\Sigma_{\mathbf{ZZ}}$  are the mean vector and covariance matrix of  $\mathbf{Z}$ ; and  $\Sigma_{\mathbf{YZ}}$  is the covariance matrix of  $\mathbf{Y}$  and  $\mathbf{Z}$ .

Given a set of evidential nodes  $\mathbf{Z}$  whose values are known to be  $\mathbf{Z} = z$ , by (21) and (22), we can derive the mean vector and covariance matrix of the unobserved nodes in  $\mathbf{Y}$ . Thus, the conditional distribution of  $\mathbf{Y}$  can be obtained. Equations (23) and (24) state that the expected values of the observed variables coincide with their observed values and their variances and covariances are null. In order to simplify the calculation, we can use an incremental method, that is, updating evidence from  $\mathbf{Z}$  one by one. Thus, we do not need to calculate the matrix inverse operation, because the matrix degenerates to a scalar. In this case,  $\Sigma_{\mathbf{YZ}}$  is a column vector and  $\Sigma_{\mathbf{ZZ}}$  is a scalar (i.e.,  $\Sigma_{\mathbf{ZZ}}^{-1} = 1/\sigma_{\mathbf{ZZ}}$ ).

If we want to give the point estimation as well as the corresponding probability intervals, we can solve the following maximum posterior distribution problem to get the point estimation, whose results are normally the conditional means:

$$\text{Maximize } f_{N(\boldsymbol{\mu}_{\mathbf{T}}, \Sigma_{\mathbf{T}})}(t_1, t_2, \dots, t_n) \prod_{a=1}^m p(v_a \mid \boldsymbol{\pi}_a) \Big|_{\mathbf{Z}}, \quad (25)$$

where  $\mathbf{Z}$  is the set of the observed variables, including those observed link flows and/or OD flows.

In summary, the specific procedure for estimating OD matrices and those unobserved link flows is given as follows.

*Step 0.* Initialize the model. According to Assumptions 3 and 4, based on prior (historical) link flows, we can determine the distribution of  $U$  and the initial matrix  $\mathbf{K}$ . Then we can obtain the initial link flows  $\mathbf{V} = (V_1, V_2, \dots, V_m) = \mathbf{K}E(U)$ . Thus the initial route choice proportion  $p_{ik}$  is calculated as follows:

$$h_a(V_a) = h_a^0 \left[ 1 + \alpha_a \left( \frac{V_a}{C_a} \right)^{\beta_a} \right], \quad (26)$$

$$c_k^i = \sum_{a \in A} h_a(V_a) \delta_{ak}^i, \quad (27)$$

$$p_{ik} = \frac{\exp(-\theta c_k^i)}{\sum_{l \in R_i} \exp(-\theta c_l^i)}, \quad (28)$$

where (26) is the link cost function, where  $h_a^0$  is the cost associated with free flow conditions,  $C_a$  is the link capacity, and

$\alpha_a$  and  $\beta_a$  are constants defining how the cost increases with traffic flow; (27) is the route cost function; (28) calculates the route choice proportion defined in (7).

*Step 1.* Solve the BN model. According to model assumptions, using the initial route choice proportion matrix  $\mathbf{P}$ , we can get the prior distribution of traffic flows (prior means and variances) using the following formulas:

$$\boldsymbol{\beta} = [(\mathbf{P}\boldsymbol{\delta})^T \mathbf{P}\boldsymbol{\delta}]^{-1} (\mathbf{P}\boldsymbol{\delta})^T, \quad (29)$$

$$E[\mathbf{V}] = \mathbf{K}E[\mathbf{U}], \quad (30)$$

$$E[\mathbf{T}] = \boldsymbol{\beta}E[\mathbf{V}] + E[\boldsymbol{\varepsilon}], \quad (31)$$

$$\mathbf{D}_\eta = \text{Diag}(\nu E[\mathbf{V}]), \quad (32)$$

$$\boldsymbol{\Sigma}_{\mathbf{V}\mathbf{V}} = \sigma_U^2 \mathbf{K}\mathbf{K}^T + \mathbf{D}_\eta, \quad (33)$$

$$\boldsymbol{\Sigma}_{\mathbf{V}\mathbf{T}} = \boldsymbol{\Sigma}_{\mathbf{V}\mathbf{V}}\boldsymbol{\beta}^T, \quad (34)$$

$$\boldsymbol{\Sigma}_{\mathbf{V}\mathbf{T}} = \boldsymbol{\Sigma}_{\mathbf{T}\mathbf{V}}, \quad (35)$$

$$\boldsymbol{\Sigma}_{\mathbf{T}\mathbf{T}} = \boldsymbol{\beta}\boldsymbol{\Sigma}_{\mathbf{V}\mathbf{V}}\boldsymbol{\beta}^T + \mathbf{D}_\varepsilon, \quad (36)$$

where (29) is for calculating the regression coefficient matrix given in (12). Equations (30) and (31) are for calculating the means of  $\mathbf{V}$  and  $\mathbf{T}$  given in (17). Equation (32) defines the diagonal variance matrix of  $\boldsymbol{\eta}$ ; that is,  $\text{Var}(\eta_a) = (E(v_a) \times \nu)^2$ , where  $\nu$  is the coefficient of variation. Equations (33) to (36) define the variance-covariance matrix in (19).

*Step 2.* Update the observed link flows, using the formulas

$$E[\mathbf{Y}\mathbf{Z} = z] = E[\mathbf{Y}] + \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Z}}\boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{Z}}^{-1}(z - E[\mathbf{Z}]), \quad (37)$$

$$\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{Z}=z} = \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Z}}\boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{Z}}^{-1}\boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{Y}}, \quad (38)$$

$$E[\mathbf{Z} | \mathbf{Z}=z] = z, \quad (39)$$

$$\boldsymbol{\Sigma}_{\mathbf{Z}|\mathbf{Z}=z} = \mathbf{0}, \quad (40)$$

$$\mathbf{T} = E[\mathbf{Y} | \mathbf{Z} = z] \Big|_{(\mathbf{Y}, \mathbf{Z}) = \mathbf{T}}, \quad (41)$$

where (37) and (38) are for updating the means and variance-covariance matrix of the unobserved variables, where  $\mathbf{Y}$  and  $\mathbf{Z}$  refer to the unobserved and observed components of  $(\mathbf{T}, \mathbf{V})$ , respectively. Equations (39) and (40) state that the expected values of the observed variables are their observed values and their variances and covariances are zero, as given in (23) and (24). Equation (41) takes the conditional means as point estimation for the OD and link flows, as the results of the maximum posterior distribution problem given in (25).

*Step 3.* Calculate the new route choice proportions. Since link flows are updated in Step 2, the route choice proportions also need to be updated. Given that the matrix  $\mathbf{T}$  is obtained by

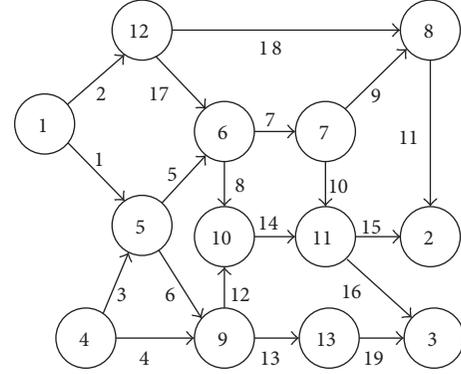


FIGURE 2: The Nguyen-Dupuis network.

(41), the new route choice proportion  $p_{ik}^*$  is calculated using the following expressions:

$$V_a = \sum_{i \in \mathbf{I}} T_i \sum_k p_{ik} \delta_{ak}^i, \quad (42)$$

$$h_a(V_a) = h_a^0 \left[ 1 + \alpha_a \left( \frac{V_a}{C_a} \right)^{\beta_a} \right], \quad (43)$$

$$c_k^i = \sum_{a \in \mathbf{A}} h_a(V_a) \delta_{ak}^i, \quad (44)$$

$$p_{ik}^* = \frac{\exp(-\theta c_k^i)}{\sum_{l \in \mathbf{R}_i} \exp(-\theta c_l^i)}, \quad (45)$$

where (42) is the conservation law equation given in (6).

*Step 4.* Test convergence. If  $\sum_{i,k} (p_{ik} - p_{ik}^*)^2 < \xi$ , where  $\xi$  is a small number to control convergence of the process, then stop the process and return the OD flow  $T_i$ , the link flow  $V_a$ , and the route choice proportion  $p_{ik}^*$ . Otherwise, continue with Step 5.

*Step 5.* Update route choice proportions and the matrix  $\mathbf{K}$ , using the expressions

$$p_{ik} = \rho p_{ik}^* + (1 - \rho) p_{ik} \quad \forall i, k, \quad (46)$$

$$\mathbf{K} = \frac{\mathbf{V}}{E(\mathbf{U})} \quad (47)$$

and go to Step 1, where (46) is for updating the route choice proportion matrix, where  $\rho$ ,  $0 < \rho < 1$ , is a relaxation factor; (47) is for updating the matrix  $\mathbf{K}$ . The values of variables  $\mathbf{V}$  are obtained by (42).

## 5. Example: The Nguyen-Dupuis Network

In this section, we illustrate the proposed methods using the well-known Nguyen-Dupuis network, shown in Figure 2. It consists of 13 nodes, 19 links, and 4 OD pairs: 1-2, 1-3, 4-2, and 4-3.

TABLE 1: Network parameters of the Nguyen-Dupuis network.

Link	$h_a^o$	$C_a$
1	7	900
2	8	700
3	9	700
4	14	900
5	5	800
6	9	600
7	5	900
8	13	500
9	5	300
10	9	400
11	10	700
12	10	700
13	9	600
14	8	700
15	9	700
16	8	700
17	7	300
18	15	700
19	11	700

The network data are shown in Table 1 and the associated parameters in (26) are assumed to be  $\alpha_a = 0.15$ ,  $\beta_a = 4$  for any link.

The assumed true OD matrices, which are used later for testing the quality of the estimation, are shown in Table 4 under the heading “True flow.” The true link flows are obtained by solving the multinomial logit assign model with parameter  $\theta = 1.0$  for the stochastic loading.

The prior information is assumed as follows: the expected value of the level of total traffic flow  $U$  and its standard deviation are  $E(U) = 50$  and  $\sigma(U) = 10$ , respectively and the initial matrix  $\mathbf{K}$  is

$$\mathbf{K} = (1.5, 0.9, 0.9, 0.7, 1.7, 0.8, 1.8, 0.1, 0.8, 1.0, 1.5, 0.3, 1.2, 0.3, 0.5, 0.8, 0.1, 0.7, 1.2)^T. \quad (48)$$

The observed link flows are assumed to be  $V_5 = 82.57$ ,  $V_7 = 87.38$ ,  $V_{10} = 48.07$ ,  $V_{13} = 58.66$ , and  $V_{18} = 37.12$ . And it is supposed that they are known in this order. Since they are observed, their values are equal to the true link flows (as shown in Table 4).

*Step 0-Step 1.* Initialize and give the prior distribution of all the variables.

Based on the prior information, we can get the prior distribution of the traffic flows. The prior means and variances are shown in the second columns of Tables 2 and 3, respectively. To simplify the calculation, in this example, the expectation and variance-covariance of  $\boldsymbol{\varepsilon}$  are assumed to be null (i.e., there is no uncertainty in the conservation law). In addition, to obtain the variance-covariance matrix  $\mathbf{D}_\eta$ , we have selected  $\nu = 0.1$  in (32).

*Step 2-Step 5.* Give the posterior distribution by updating the observed link flows one by one.

TABLE 2: Point estimation of traffic flows after updating observed link flows one by one.

Item	0	5	7	10	13	18
$T_1$	37.16	39.02	39.05	39.86	40.11	39.88
$T_2$	82.88	77.95	78.32	78.63	80.51	81.22
$T_3$	68.37	59.64	59.94	59.23	60.85	59.95
$T_4$	12.68	22.48	22.29	23.28	19.54	19.06
$V_1$	75.00	74.64	74.92	75.26	76.98	77.55
$V_2$	45.00	42.36	42.48	43.28	43.69	43.61
$V_3$	45.00	46.30	46.32	45.73	46.22	46.03
$V_4$	35.00	35.82	35.92	36.78	34.14	32.80
$V_5$	85.00	<b>82.57</b>	<b>82.57</b>	<b>82.57</b>	<b>82.57</b>	<b>82.57</b>
$V_6$	40.00	38.37	38.62	38.37	40.69	41.17
$V_7$	90.00	87.24	<b>87.38</b>	<b>87.38</b>	<b>87.38</b>	<b>87.38</b>
$V_8$	5.00	1.43	1.45	1.55	1.51	1.54
$V_9$	40.00	39.38	39.40	<b>39.31</b>	<b>39.31</b>	<b>39.31</b>
$V_{10}$	50.00	47.86	47.92	<b>48.07</b>	<b>48.07</b>	<b>48.07</b>
$V_{11}$	75.00	75.64	75.73	76.24	76.58	<b>76.43</b>
$V_{12}$	15.00	15.14	15.42	15.44	16.20	15.43
$V_{13}$	60.00	59.05	59.12	59.71	<b>58.66</b>	<b>58.66</b>
$V_{14}$	15.00	16.57	16.87	16.99	17.71	<b>16.84</b>
$V_{15}$	25.00	23.06	23.31	22.89	24.42	<b>23.57</b>
$V_{16}$	40.00	41.37	41.49	42.22	<b>41.34</b>	<b>41.34</b>
$V_{17}$	5.00	6.10	6.16	6.29	6.40	6.51
$V_{18}$	35.00	36.26	36.32	36.40	37.29	<b>37.12</b>
$V_{19}$	60.00	59.05	59.12	59.71	<b>58.66</b>	<b>58.66</b>

Table 2 shows how the means of the traffic flows changed after updating the observed link flows one by one. After each update, the point estimation of link flows  $\{V_1, V_2, \dots, V_{19}\}$  and OD flows  $\{T_1, T_2, \dots, T_4\}$  is provided. It can be seen that once  $V_7$  and  $V_{10}$  become known and updated, the point estimation of  $V_9$  in Table 2 remains unchanged and its variance in Table 3 becomes zero (boldfaced in the table). Because, due to the flow conservation in node 6, once  $V_7$  and  $V_{10}$  become known,  $V_9$  becomes known. Similarly,  $V_{19}$  becomes known once  $V_{13}$  is given;  $V_{16}$  becomes known once  $V_{19}$  is given;  $V_{11}$  becomes known once  $V_9$  and  $V_{18}$  are given;  $V_{15}$  becomes known once  $V_{11}$  is given;  $V_{14}$  becomes known once  $V_{10}$ ,  $V_{15}$ , and  $V_{16}$  are given. In other words, due to the conservation law, the observed link flows are known in this order:  $V_5 = 82.57$ ,  $V_7 = 87.38$ ,  $V_{10} = 48.07$ ,  $V_9 = 39.31$ ,  $V_{13} = 58.66$ ,  $V_{19} = 58.66$ ,  $V_{16} = 41.24$ ,  $V_{18} = 37.12$ ,  $V_{11} = 76.43$ ,  $V_{15} = 23.57$ , and  $V_{14} = 16.84$ .

Table 3 shows how the variances of the traffic flows changed after updating the observed link flows one by one. Note that after some link flows (including the observed link flows and those derived from the observed link flows and the conservation laws) are known, their means remain constant and their variances become null (boldfaced in the tables). And normally the variances of the unknown variables (OD flows and those unobserved link flows) decrease with each update. Note that the smaller the variances, the higher the precision of the estimation. So after a series of update, the estimation becomes more accurate. This derives a method to

TABLE 3: Variances of traffic flows after updating observed link flows one by one.

Item	0	5	7	10	13	18
$T_1$	65.81	25.98	20.88	16.94	14.35	2.93
$T_2$	294.56	54.98	29.91	20.46	15.66	15.66
$T_3$	203.36	22.72	7.59	5.41	5.40	3.30
$T_4$	18.56	16.98	15.59	14.10	3.24	3.16
$V_1$	281.25	100.33	81.13	69.99	67.48	65.57
$V_2$	101.25	32.31	26.08	23.15	21.73	20.73
$V_3$	101.25	38.61	31.01	25.85	24.33	23.10
$V_4$	61.25	23.11	18.65	16.72	13.27	11.73
$V_5$	361.25	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
$V_6$	80.00	26.51	21.55	18.19	18.85	18.48
$V_7$	405.00	137.07	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
$V_8$	1.25	0.04	0.03	0.03	0.03	0.03
$V_9$	80.00	27.93	22.44	<b>0</b>	<b>0</b>	<b>0</b>
$V_{10}$	125.00	41.25	33.19	<b>0</b>	<b>0</b>	<b>0</b>
$V_{11}$	281.25	103.05	82.88	71.83	66.79	<b>0</b>
$V_{12}$	11.25	4.13	3.44	2.95	2.99	2.60
$V_{13}$	180.00	62.81	50.52	44.05	<b>0</b>	<b>0</b>
$V_{14}$	11.25	4.94	4.11	3.57	3.57	<b>0</b>
$V_{15}$	31.25	9.58	7.85	6.47	6.79	<b>0</b>
$V_{16}$	80.00	30.82	24.88	22.03	<b>0</b>	<b>0</b>
$V_{17}$	1.25	0.67	0.55	0.50	0.47	0.46
$V_{18}$	61.25	23.68	19.07	16.91	15.83	<b>0</b>
$V_{19}$	180.00	62.81	50.52	44.05	<b>0</b>	<b>0</b>

determine how many links and what links need to be observed when estimating traffic flows by the Bayesian network model, that is, the network sensor location problem (NLSP) [28]. Note that the variances updating equation (22) has no relevance with the values of the observed link flows. So we can solve NLSP without observing any link. First, by the Bayesian network model, we can get the prior distribution of all the variables shown in (19). Next, we can take the link which can reduce the variance of the OD flows maximally by updating as the first observed link. Then we update the variances of traffic flows, determine the second observed link, and iterate until the variances decrease to meet the requirement of the estimation precision or until the budget exceeds the constraint.

By the updated means (point estimation) and variances, we can obtain the posterior distribution of OD flows and those unobserved link flows. Figure 3 illustrates how the marginal densities of OD flows and those unobserved link flows evolve from their initial form to their final form (boldfaced) by updating the observed link flows one by one. It can be seen that the variances of the unknown variables are normally decreasing with each update.

In summary, according to Tables 2 and 3 and Figure 3, using the proposed Bayesian network method, after some variables are observed, the means of these observed remain constant and their variances become null. For the remaining variables (those unobserved), their variances are normally

TABLE 4: The true flow and the point estimation from the proposed method.

Item	True flow	Proposed method	Relative error (%)
$T_1$	<b>40</b>	<b>39.88</b>	<b>0.30</b>
$T_2$	<b>80</b>	<b>81.22</b>	<b>1.53</b>
$T_3$	<b>60</b>	<b>59.95</b>	<b>0.08</b>
$T_4$	<b>20</b>	<b>19.06</b>	<b>4.70</b>
$V_1$	<b>76.64</b>	77.55	<b>1.19</b>
$V_2$	<b>43.36</b>	<b>43.61</b>	<b>0.58</b>
$V_3$	<b>46.13</b>	<b>46.03</b>	<b>0.22</b>
$V_4$	<b>33.87</b>	<b>32.80</b>	<b>3.16</b>
$V_5$	82.57	82.57	0
$V_6$	<b>40.21</b>	<b>41.17</b>	<b>2.39</b>
$V_7$	87.38	87.38	0
$V_8$	<b>1.42</b>	<b>1.54</b>	<b>8.45</b>
$V_9$	39.31	39.31	0
$V_{10}$	48.07	48.07	0
$V_{11}$	76.43	76.43	0
$V_{12}$	<b>15.41</b>	<b>15.43</b>	<b>0.13</b>
$V_{13}$	58.66	58.66	0
$V_{14}$	16.84	16.84	0
$V_{15}$	23.57	23.57	0
$V_{16}$	41.34	41.34	0
$V_{17}$	<b>6.24</b>	<b>6.51</b>	<b>4.33</b>
$V_{18}$	37.12	37.12	0
$V_{19}$	58.66	58.66	0

decreasing with each update. The proposed method can provide a control of the conservation law as well. The final forms (boldfaced) in Figure 3 are the posterior densities of OD flows and those unobserved link flows. These posterior densities supply complete statistical information about the unknown variables. By these posterior densities, we can provide the point estimation as well as the corresponding probability intervals.

In order to test the quality of the estimation, Table 4 compares the true flow and the point estimation of the proposed BN model. Because  $V_5$ ,  $V_7$ ,  $V_9$ ,  $V_{10}$ ,  $V_{11}$ ,  $V_{13}$ ,  $V_{14}$ ,  $V_{15}$ ,  $V_{16}$ ,  $V_{18}$ , and  $V_{19}$  are observed, their values are equal to the true flow. And studying the OD pairs and those unobserved links (boldfaced in the table), it can be seen that the estimation and true flow are basically the same. The relative errors are all very small and the maximum relative error value of the OD flows estimation is only 4.70%. This illustrates that the proposed BN model has a high accuracy.

## 6. Conclusions

In this paper, we use a Bayesian network model to estimate origin-destination matrices based on prior link flows and a set of observed link flows. Normally, large amounts of historical link flows are stored in the cities' transportation system. Compared with an outdated or subjectively guessed prior OD matrix, prior link flows are more accurate as they are obtained by traffic detectors or manual investigation. The proposed

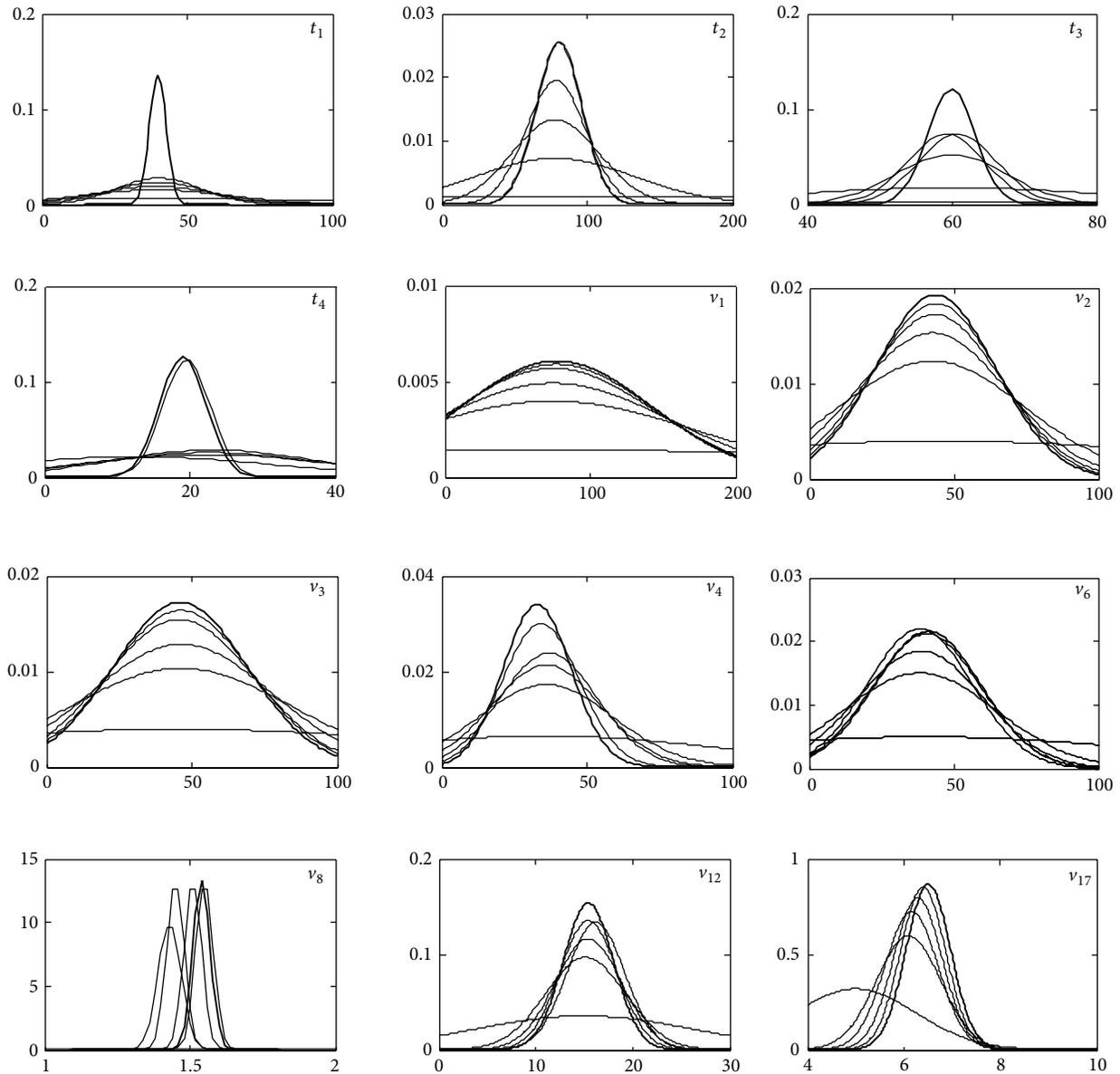


FIGURE 3: Conditional distributions of the OD flows and the unobserved link flows.

Bayesian network model can make use of these historical link flows and also consider the level of total traffic flow, which is really useful for many real-time traffic applications, especially in the ITS.

Using the Bayesian network model and updating the observed variables (including the observed link flows and those derived from the observed link flows and the conservation law) can modify the means and reduce the variances of the remaining variables. These updated means and variances allow us to obtain the posterior distribution of the unobserved variables based on those observed. Thus, the methods can provide not only point estimation but also the corresponding probability intervals. In addition, an incremental procedure is developed for solving the Bayesian network model without the intensive computation of matrix

inversion, which can make this model apply easily in large-scale networks.

Moreover, in this paper, a normal distribution for traffic flows is assumed. It is reasonable because these random variables are the sum of a great number of independent Bernoulli experiments in which the users decide where to travel and which routes to choose. For future research, it is worthwhile to relax the normal distribution assumption.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research is supported by the National Natural Science Foundation of China (no. 51178110 and no. 51378119) and Graduate Innovation Project of Jiangsu Province (no. CXZZ12\_0113). Comments provided by anonymous referees are much appreciated.

## References

- [1] M. Ben-Akiva, M. Bierlaire, J. Bottom, H. Koutsopoulos, and R. Mishalani, "Development of a route guidance generation system for real-time application," in *Proceedings of the IFAC Transportation Systems Conference*, Beijing, China, 1997.
- [2] H. S. Mahmassani, "Dynamic network traffic assignment and simulation methodology for advanced system management applications," *Network and Spatial Economics*, vol. 1, no. 3-4, pp. 267–292, 2001.
- [3] M. G. Bell, "The estimation of origin-destination matrix from traffic counts," *Transportation Science*, vol. 17, no. 2, pp. 198–217, 1983.
- [4] E. Cascetta and S. Nguyen, "A unified framework for estimating or updating origin/destination matrices from traffic counts," *Transportation Research B: Methodological*, vol. 22, no. 6, pp. 437–455, 1988.
- [5] H. Yang, "Heuristic algorithms for the bilevel origin-destination matrix estimation problem," *Transportation Research B: Methodological*, vol. 29, no. 4, pp. 231–242, 1995.
- [6] J. Doblas and F. G. Benitez, "An approach to estimating and updating origin-destination matrices based upon traffic counts preserving the prior structure of a survey matrix," *Transportation Research B: Methodological*, vol. 39, no. 7, pp. 565–591, 2005.
- [7] Y. Nie and H. M. Zhang, "A variational inequality formulation for inferring dynamic origin-destination travel demands," *Transportation Research B: Methodological*, vol. 42, no. 7-8, pp. 635–662, 2008.
- [8] Y. M. Nie and H. M. Zhang, "A relaxation approach for estimating origin-destination trip tables," *Networks and Spatial Economics*, vol. 10, no. 1, pp. 147–172, 2010.
- [9] W. Shen and L. Wynter, "A new one-level convex optimization approach for estimating origin-destination demand," *Transportation Research B: Methodological*, vol. 46, no. 10, pp. 1535–1555, 2012.
- [10] M. Carey and R. Revelli, "Constrained estimation of direct demand functions and trip matrices," *Transportation Science*, vol. 20, no. 3, pp. 143–152, 1986.
- [11] M. G. H. Bell, "The estimation of origin-destination matrices by constrained generalised least squares," *Transportation Research B: Methodological*, vol. 25, no. 1, pp. 13–22, 1991.
- [12] K. Parry and M. L. Hazelton, "Estimation of origin-destination matrices from link counts and sporadic routing data," *Transportation Research B: Methodological*, vol. 46, no. 1, pp. 175–188, 2012.
- [13] H. J. van Zuylen and L. G. Willumsen, "The most likely trip matrix estimated from traffic counts," *Transportation Research B: Methodological*, vol. 14, no. 3, pp. 281–293, 1980.
- [14] C. Xie, K. M. Kockelman, and S. T. Waller, "A maximum entropy-least squares estimator for elastic origin-destination trip matrix estimation," *Transportation Research B: Methodological*, vol. 45, no. 9, pp. 1465–1482, 2011.
- [15] H. P. Lo, N. Zhang, and W. H. K. Lam, "Estimation of an origin-destination matrix with random link choice proportions: a statistical approach," *Transportation Research B: Methodological*, vol. 30, no. 4, pp. 309–324, 1996.
- [16] Y. Vardi, "Network tomography: estimating source-destination traffic intensities from link data," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 365–377, 1996.
- [17] M. L. Hazelton, "Estimation of origin-destination matrices from link flows on uncongested networks," *Transportation Research B: Methodological*, vol. 34, no. 7, pp. 549–566, 2000.
- [18] M. J. Maher, "Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach," *Transportation Research B: Methodological*, vol. 17, no. 6, pp. 435–447, 1983.
- [19] B. Li, "Bayesian inference for origin-destination matrices of transport networks using the em algorithm," *Technometrics*, vol. 47, no. 4, pp. 399–408, 2005.
- [20] M. L. Hazelton, "Statistical inference for time varying origin-destination matrices," *Transportation Research B: Methodological*, vol. 42, no. 6, pp. 542–552, 2008.
- [21] M. L. Hazelton, "Statistical inference for transit system origin-destination matrices," *Technometrics*, vol. 52, no. 2, pp. 221–230, 2010.
- [22] C. Tebaldi and M. West, "Bayesian inference on network traffic using link count data," *Journal of the American Statistical Association*, vol. 93, no. 442, pp. 557–573, 1998.
- [23] S. Sun, C. Zhang, and G. Yu, "A Bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124–133, 2006.
- [24] E. Castillo, J. M. Menéndez, and S. Sánchez-Cambronero, "Predicting traffic flow using Bayesian networks," *Transportation Research B: Methodological*, vol. 42, no. 5, pp. 482–509, 2008.
- [25] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufman, San Mateo, Calif, USA, 1988.
- [26] F. V. Jensen, *Bayesian Networks and Decision Graphs*, Springer, 2001.
- [27] R. Shachter and C. Kenley, "Gaussian influence diagrams," *Management Science*, vol. 35, no. 5, pp. 527–550, 1989.
- [28] S. L. Zhu, L. Cheng, Z. M. Chu, C. Anthony, and J. X. Chen, "Identification of network sensor locations for traffic flow estimation," accepted by *Journal of the Transportation Research Board*.

*Research Article*

# **A Conceptual Architecture for Adaptive Human-Computer Interface of a PT Operation Platform Based on Context-Awareness**

**Qing Xue, Xuan Han, Mingrui Li, and Minxia Liu**

*Beijing Institute of Technology, School of Mechanical Engineering, Beijing 100081, China*

Correspondence should be addressed to Qing Xue; xueqing@bit.edu.cn

Received 31 December 2013; Revised 14 February 2014; Accepted 4 March 2014; Published 3 April 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Qing Xue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a conceptual architecture for adaptive human-computer interface of a PT operation platform based on context-awareness. This architecture will form the basis of design for such an interface. This paper describes components, key technologies, and working principles of the architecture. The critical contents covered context information modeling, processing, relationship establishing between contexts and interface design knowledge by use of adaptive knowledge reasoning, and visualization implementing of adaptive interface with the aid of interface tools technology.

## **1. Introduction to AHCI of PT Operation Platform Based on Context-Awareness**

Adaptive systems are usually embedded in larger human-machine systems such as the autopilot, warning, or navigation systems which could enhance human-machine system performance [1]. Context-awareness is critical for military agencies to benefit operational effectiveness by facilitating the planning process, improving the quality and timeliness of decisions, and providing better feedback regarding the strategic consequences of military actions based on knowledge technologies and visualization components [2]. But the HCI (human-computer interface) for military combat vehicles now lacks adaptability, and its intelligent implementation mechanisms cannot be suitable for all kinds of special transportation vehicles.

As humans and computers generally undertake their respective part in intelligent transportation vehicle, human is in charge of the supervision of the overall task at the higher level [3]. But as the tasks allocated to the computer increase, the heavy cognitive load and defective context-awareness caused inevitable errors and tedious work during operation process. Adaptive interface is useful to support the operators in complex dynamic task controls [4]. AHCI (adaptive human-computer interface) is an interface which

is supposed to adapt its interaction contents, information processing modes, and behavior patterns automatically to meet changing task requirements and user characteristics at any time considering operator abilities, workload variations, and skill levels [5].

PT operation platform is a real-time interactive platform on military intelligent transportation vehicle, whose interface directs impacts on huge amounts of information transmission and combat efficiency in the information warfare. PT's interface is influenced by tasks, devices, environments, and other contextual factors that determine the modes of task execution and the patterns of interaction. The current situation is that the HCI of PT is limited by mobility, computing ability, input/output modes, and some other factors; although it can support the task execution, it still lacks user friendliness and autonomy. To address these challenging HCI demands, it is possible to understand the requests of states and users well enough to adapt the interface components, content, structure, and form, in terms of the dynamic changed situation, as well as to provide the operator with the necessary data and information [4, 6]. Therefore, studying new human-computer interaction modes and designing AHCI based on PT's characteristics become a new important subject which would be capable of detecting and responding to changing contexts of the user and the task in the PT operation platform.

In order to design and develop AHCI of the PT operation platform on the basis of mission analysis and reasonable human-computer function assignment, a conceptual architecture was established on general AHCI theory in this paper. Technologies of context-awareness were used in adaptive mechanism in response to dynamic context changes. The favorable context-modeling technology for the context types and the methods of adaptive knowledge reasoning were discussed according to the practical application of PT, by which a mapping relationship between context information and interface visualization could be generated. And an AHCI of the PT can be realized based on the visualization modes and interface design tools, which is supposed to provide information and operation services appropriately for different users, stages, and tasks automatically in time.

The ultimate goal is to achieve visualization of battlefield contexts and customization of user needs and the tools of interface development to enhance the operational efficiency of the PT operation platform.

## 2. A Conceptual Architecture for AHCI of the PT Operation Platform Based on Context-Awareness

Context-awareness is the concept of describing the performance of a system and getting a task-specific understanding during the complex operation [7]. It was first presented in connection with pilot performance in air-to-air combat and the ability of commercial airline pilots to fly in difficult air traffic condition [8]. Similarly, PT is also a complex dynamic control system in which context variables change over time.

For the PT operation platform, the main goals and functions of AHCI are shown in Table 1.

According to the specific application environment of the PT operation platform, the AHCI conceptual architecture based on context-awareness is shown in Figure 1, which is mainly constituted of the device layer, the adaptive control layer, the interface configuration layer, and the user interface layer.

*2.1. Device Layer.* The function of this layer is to perceive atomic context information from external and internal sources. The first kind of context information is external context which includes postures, target parameters, working conditions, battlefield environment data, and external task commands that were obtained from physical sensors or peripheral devices, such as GPS, radar, electrooptical device, and command and control systems. Sensor data is simply sampled and packaged into a digital signal that can be transferred to a computer for further processing [9]. The second kind of context comes from internal equipment by recording the system output, device current states, and interface display situation. The last kind of context information is interaction history which contains user input to the system and the feedback from the system.

Data collected via various sources can be used to make assessments of the system state, the environment, the task,

and the user. Adaptive decisions will be generated by use of these assessments to decide which adaptations to select [1].

*2.2. Adaptive Control Layer.* There are two functions in this layer.

The first function is context information processing. A rough knowledge representation can be established by classified atomic contexts acquired from device layer. Atomic contexts can be refined into composite contexts after consistency test and fusion, which will be imported to reasoning engine consisting of rules and learning methods.

The second function is to make adaptive decisions by reasoning. The reasoning process may contain more inference activities that are triggered by degrees, after which mapping relationships between the context space and the user interface space can be established. Finally, adaptive decision-making that is suitable for the current situation will be put forward and request the interface rendering engine to generate the interface.

In the AHCI knowledge base, the context base stores context knowledge. Former cases and inference rules are stored in rule base.

*2.3. Interface Configuration Layer.* According to adaptive decisions, interface rendering engine extracts the interface elements and templates from the interface elements base to assemble a user interface which is suitable for the current users, tasks, and environment. Main interface attributes involve the interactive patterns, contents, display styles, and so forth. This layer contains an important mechanism to realize the adaptive conversion of interface configuration.

*2.4. User Interface Layer.* The interface in this layer was assembled by interface configuration layer which realized adaptive information visualization and adaptive task operation. As the new context, the interaction log between user and the system will be sent back to device layer in supporting next adaptive process.

## 3. AHCI Working Principle of the PT Operation Platform Based on Context-Awareness

### 3.1. Context Information Processing

*3.1.1. Contexts of the PT Operation Platform.* Context is any instantaneous, detectable, and relevant property related to future device actions that describes the current situation on an abstract level [9, 10]. It usually comes from the end users of the interactive system, the hardware and software computing platforms with which their interactive tasks are carried out, and the physical environment where they are working [11].

Context information modeling is an important step when designing AHCI based on context-awareness. After the rough refining, the high level composite contexts can be generated for the further adaptive decision-making. The contexts of PT are divided into the following categories as shown in Table 2.

TABLE 1: Goals and functions of AHCI for the PT operation platform.

Main goals	Description	Examples	Description of function
Improve usability	Make system convenient and efficient	Default setting	Improve the understanding ability of the system, avoid repeated execution (default based on the highest operation frequency or latest operation record)
Simplify the operation process	Combine process steps, omit redundant operations	Simplifying tasks	Simplify the task steps by batch processing and macro command, execute tasks automatically without user's confirmation
Reduce operation complexity	Reduce operation process and cognitive load	Simplifying the interfaces	Hide the function components which are not in need
Display information appropriately	Provide needed information timely and properly	Guiding	Help operators to accomplish mission tasks
		Error correction	Indicate errors and correct them in time
		Information push server	Provide information autonomously according to the requirements
		Status tracking	Reflect situation in real time
Support user needs diversification	Meet the demands of different users	Customizing interface	Transform component configurations according to user characteristics

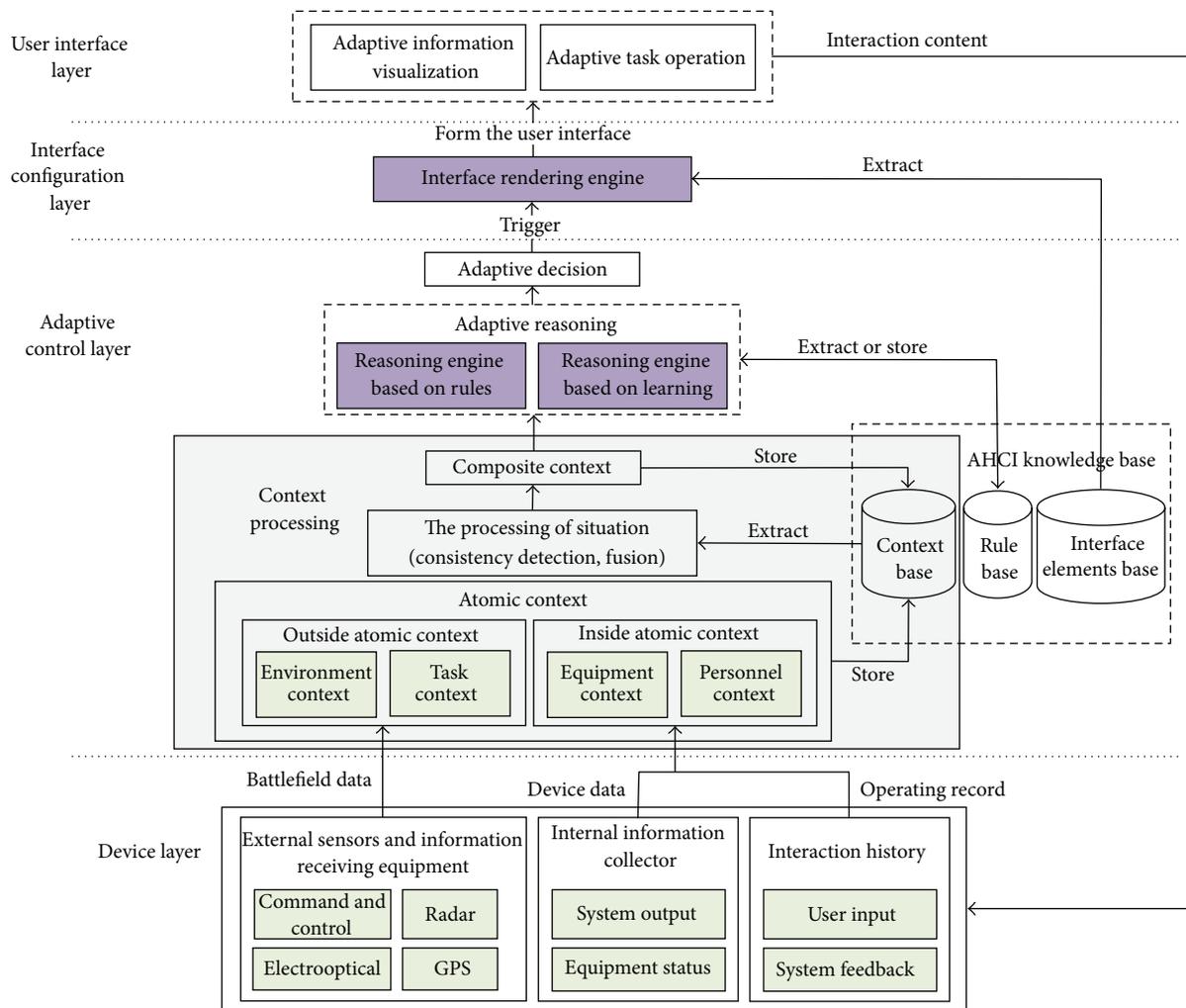


FIGURE 1: Conceptual architecture for AHCI of the PT operation platform based on context-awareness.

TABLE 2: Categories of the PT's contexts.

Context categories	Explanation
User context	Describe the different users such as the conductor and the driver
Equipment context	Describe the equipment states, both hardware and software
Environment context	Describe the physical environment, such as temperature, humidity, and visibility
Time and space context	Describe the time, location, and so forth
Social context	Describe the external information communication between collaborations
Task context	Describe the target, task processing, and operation flow
Display and control context	Describe the current layout, color, operation mode, and so forth

TABLE 3: Characteristics of PT's contexts.

Characteristics	Explanation
Concurrency	There are concurrent features among internal and external tasks.
Wide distribution	In the battlefield, context sources are widely distributed.
Dynamic nature	Context varies in terms of time, space, and interactive mode that affect the operation behavior.
Sociality	External communication during the task execution will have an impact on other contexts.

Equipment, environment, time, and space contexts can be obtained directly from sensors. User context is obtained directly from user's input or interaction records. Social context and task context are acquired during the process of the mission. According to the mission analysis and functional assignation including object, goal, logical function, and task scheduling, contexts above at a certain moment would be combined with the current display and control context to be reasoned adaptively to generate causal relationships between context and display control interface.

Due to the particularity application environment, these contexts have the characteristics shown in Table 3.

Because of the particularity and limitations, real-time context knowledge acquisition, storage, modeling, integration, and reasoning for the PT operation platform are very important and complex.

*3.1.2. Context-Modeling Method for AHCI of the PT.* Context information both accurate and fuzzy should be used for adaptive decision-making latterly in implementation of AHCI. The first step is to establish a context model. Ontology and logical modeling were presented to establish the context model in this paper.

*(1) Context-Modeling Based on Ontology.* Ontology provides a foundation for semantic interpretation and information fusion processes, while facilitating information search and retrieval, knowledge elicitation, knowledge modeling, and knowledge representation [2].

Ontology modeling presents semantic concepts and attributes of context entities, as well as the relationships among them by using ontology theory. Ontology modeling tool *Protégé* can test consistency of context information with inner adaptive reasoning method. But the context interpretation based on ontology requires correct and complete information, so it is hard to express inaccurate contexts with this approach.

Building ontology model for accurate and complete context information of PT's interface, the ontology is the specification of all conceptual elements in every combat phase that affect PT's display and control. The application is mainly involved in domain ontology and task ontology. The former describes the static specified domain knowledge and gives the concepts, relationships, activities, characteristics, and laws of domain entities; the latter describes hierarchical and interdependent relationships of task status concepts which reflect the dynamic behaviors in task changes.

Taking the fighting task as an example, the domain ontology model was built according to task analysis as shown in Figure 2. Context ontology and its relationships in combat implementation stage were set up in domain model.

*(2) Context-Modeling Based on Logic.* Most context information of PT is naturally uncertain and fuzzy because of sensor data missing or semantics confusion; context model should be designed in a way that allows the representation of uncertainty [12].

Based on logical method, context can be defined by a series of facts, expressions, and rules to support contexts' reasoning processing. Logical rules mainly define constraints conditions so that new facts or expressions could be deduced and inferred from existing facts or expressions. The lists of facts and expressions could be added, deleted, and updated. With solid mathematical foundation, the modeling based on logic has obvious advantages on formalization and representation of uncertainty context.

For PT, context-modeling based on logic also can be used in subsequent adaptive reasoning to generate mapping relationships between contexts and interface mode.

*3.2. Adaptive Interface Knowledge Reasoning.* The AHCI must be decision centric which is able to advise the user when to take over control of the system and when to supervise control tasks [4].

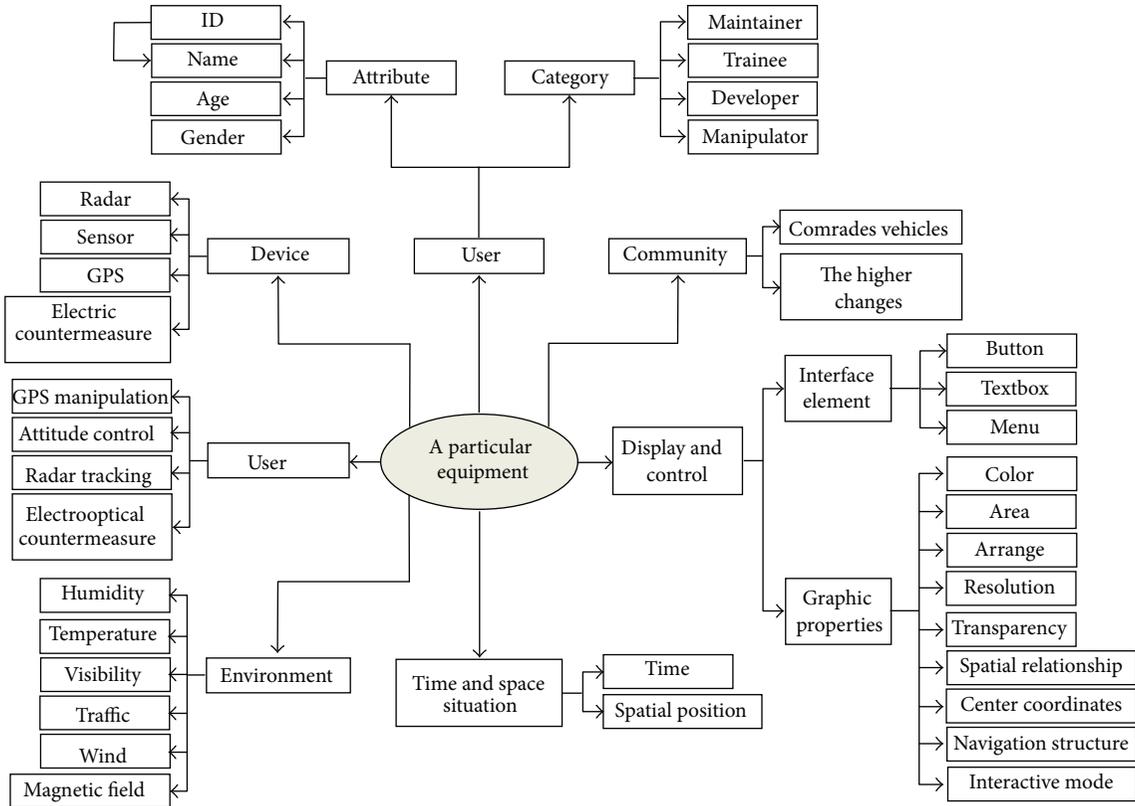


FIGURE 2: PT’s domain ontology.

Adaptive context reasoning is to provide new and relevant information which is derived from context data sources to be used by the given application or the user [13, 14]. It is to identify composite contexts and then match adaptive responses to the current situation through the reasoning engine according to certain logical rules. Adaptive reasoning should give user effective responses, maximum accuracy, and efficiency task completion, as well as the most optimal display mode to minimize the cognitive and operation fatigue.

There are two main adaptive aspects for the AHCI of the PT operation platform: adaptive for information visualization and adaptive for user task operation. With regard to different categories of PT’s contexts and different adaptive goals, adaptive reasoning technology should be carried out not only for display related knowledge but also for task operation related knowledge. In this paper, Ontology-Based Reasoning and Case-Based Reasoning were proposed to be used respectively or together according to their advantages in decision-making.

(1) *Ontology-Based Reasoning.* Ontology-Based Reasoning uses ontology axiom to realize the collision detection, express optimization, ontology integration, and validation by limitation of ontology model and its own description language. So Ontology-Based Reasoning would be used to complete preliminary reasoning of composite context. As Ontology-Based Reasoning requires custom logical rules, all domain

ontology knowledge and training models have been defined before; in high level reasoning, other methods would be used on the basis of Ontology-Based Reasoning’s preliminary result.

(2) *Case-Based Reasoning.* CBR (Case-Based Reasoning) is a method that uses previous cases as reference to infer possible cases. The procedure is matching the current context with history context to predict possible decisions and to optimize rules and inference mechanism through the summary and analysis of cases. As CBR can deal with complex and uncertain knowledge, it would be used on focused of extracting and analyzing cases automatically. For PT’s contexts which are incomplete and indistinguishable, it is hard to build exact relationships one by one between context and interface mode. Rough sets are a kind of CBR method that can realize adaptive knowledge simplification and reasoning on PT’s task operation adaptation to get a mapping from the context space (CONS) to user interface space (UIS), represented as IF CONS, THEN UIS.

Adaptive knowledge reasoning for information visualization on PT is to realize the optimal interface mode by matching the most appropriate interface elements to compose an interface adaptively. The interface property is a tuple of the form  $UIS = \langle L, N, S, A \rangle$ , in which L is the layout arrangement, N is the hierarchy structure of navigation, hyperlinks, and menu, S is the interactive style, and A is the set of display attributes that includes color and size [15].

**3.3. The Working Mode of Adaptive Interface Visualization.** Interface visualization is an expression technology by which information data is converted to intuitive understanding image and graphics shown on the interface in order to provide commander a clear understanding of the current state during mission accomplishment [16]. It is a mapping process from data form to the visual form.

AHCI visualization of the PT is the working process that interface rendering engine is triggered by adaptive decisions to present the visualization mapping results using the interface elements extracted from interface elements base. The visualization procedure includes data transformation, view transformation, content adaptation, dialog adaptation, task allocation, and decision support as well as the transformation between interfaces [3].

There are two interface variant ways in visualization modes which are dynamic interface behaviors by “changing” the original interface [17]. For static variation, the different interface templates that were designed in advance are waiting to be called and selected by user to display on the interface. For dynamic variation, the system judges the situation automatically by analyzing the contexts and matches the most appropriate interface elements and features to generate specific alternative mode.

The antecedents of visualization mapping are adaptive decision rules, while the consequents are corresponding interface mode, whose element attributes can be represented in set of Icon =  $\langle \text{tag, graphic attributes} \rangle$ , in which tag is the abstract symbol of visual objects, such as spot, line, and face, while graphic attributes include tag’s center coordinates  $(x, y)$ , area  $S$ , arrangement, resolution, transparency (0–100%), and the spatial relations among multi-icons. The attribute configurations and values will be assigned under various contexts’ triggers.

## 4. Interface Tools Technology of the PT’s AHCI

In order to implement adaptive interface and to transplant the interface into other similar vehicles easily, it is necessary to explore the adaptive interface tools technology.

**4.1. The Model-Based User Interface Technology.** Interface model is an important part in AHCI knowledge base. It describes different aspects of the interface by declaration. It consists of some typical models, such as task model, data model, display model, and layout model. In fact, the development of the model-based user interface means how to describe the whole interface in a better way by these models. Departing from users’ perspective, the interface is divided into two parts, visibility and usability, while PT’s display-control system is also divided into two parts, interface and logic, which guarantees the appearance of the adaptive interface.

**4.2. The XML-Based Interface Management.** The XML-based interface management gives the solution of converting the display and the layout model into the specific interface.

The model-based user interface technology obtains partial and integral descriptions of the interface display modes which are established by the interface description language XML. The interface is obtained after analyzing the XML documents by use of interface elements and interface templates. Interface templates are about the overall description of interface, which include the overall interface style and layout on a given type of PT platform. By invoking the templates, the decision result could be displayed on the interface under the adaptive interface architecture.

**4.3. The Development of PT’s Interface Elements Base.** The PT’s interface elements base is important in the embedded graphics system whose development is a necessary way to bring out the interface tools technology. After generalizing and arranging common interface elements related to PT operating, the interface elements base will be created by development tool Wind ML (Wind Media Library). Actually WINDOW is a basic component of Wind ML. Departing from the WINDOW, researching on *Display* and *Event*, the needed elements will be extracted from the PT’s interface applying the object-oriented idea and then the element classes will be established. Moreover, the common property and the operation object will also be extracted to be packaged into the element classes. Based on the above, the elements base that has better portability and conformity will be guaranteed.

## 5. Conclusion

Nowadays, we are concerned with dynamic contexts management in interface design knowledge for real-time human-computer interaction in the military intelligent transportation vehicle.

This paper put forward the idea of AHCI aiming to improve task execution efficiency and reduce the cognitive load and the operation complexity of operators on PT. An AHCI conceptual architecture for the PT operation platform based on context-awareness was presented whose working mode was divided into three steps. Firstly, PT’s context model was built based on ontology and logical method. Secondly, preliminary test and fusion of the context data would be implemented with the ontology reasoning method, while mapping relationship between context space and interface space could be established based on the CBR reasoning. Finally, adaptive interface visualization was supposed to be realized with help of interface design tools according to the adaptive decision. The structure proposed in this paper makes it possible to quantize all the information and thus enables using adaptive analyzing and visualization mechanisms in the PT military combat vehicle.

As context-awareness is critical on the modern battlefield, a large amount of intelligence information should be collected for decision-making process. In many cases, because of the lack of reasonable tools and methods to process context information, analysis and decision-making may be decelerated [18]. On the basis of the existing research, we will continue to study the analytical and visualization methods in depth

to improve the decision-making process of AHCI on such special transportation vehicle and to expand its generality.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgment

The work was supported in part by the preresearch project. This paper only reflects the authors' view; funding agencies are not liable for any use that may be made of the information contained herein.

### References

- [1] K. M. Feigh, M. C. Dorneich, and C. C. Hayes, "Toward a characterization of adaptive systems a framework for researchers and system designers," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 54, no. 6, pp. 1008–1024, 2012.
- [2] P. R. Smart, A. Russell, N. R. Shadbolt, M. C. Shraefel, and L. A. Carr, "AKTiveSA: a technical demonstrator system for enhanced situation awareness," *Computer Journal*, vol. 50, no. 6, pp. 703–716, 2007.
- [3] G. G. Yen and D. Acay, "Adaptive user interfaces in complex supervisory tasks," *ISA Transactions*, vol. 48, no. 2, pp. 196–205, 2009.
- [4] E. Letsu-Dake and C. A. Ntuen, "A case study of experimental evaluation of adaptive interfaces," *International Journal of Industrial Ergonomics*, vol. 40, no. 1, pp. 34–40, 2010.
- [5] E. Letsu-Dake and C. A. Ntuen, "A conceptual model for designing adaptive human-computer interfaces using the living systems theory," *Systems Research and Behavioral Science*, vol. 26, no. 1, pp. 15–27, 2009.
- [6] W. W. Noah and S. M. Halpin, "Adaptive user interfaces for planning and decision aids in C-3I systems," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 16, no. 6, pp. 909–918, 1986.
- [7] Y. J. Kim and C. M. Hoffmann, "Enhanced battlefield visualization for situation awareness," *Computers and Graphics (Pergamon)*, vol. 27, no. 6, pp. 873–885, 2003.
- [8] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: a fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.
- [9] A. Schmidt, M. Beigl, and H.-W. Gellersen, "There is more to context than location," *Computers and Graphics (Pergamon)*, vol. 23, no. 6, pp. 893–901, 1999.
- [10] S. S. Yau, D. Huang, H. Gong, and Y. Yao, "Support for situation awareness in trustworthy ubiquitous computing application software," *Software—Practice and Experience*, vol. 36, no. 9, pp. 893–921, 2006.
- [11] G. Calvary, J. Coutaz, D. Thevenin, Q. Limbourg, L. Bouillon, and J. Vanderdonckt, "A unifying reference framework for multi-target user interfaces," *Interacting with Computers*, vol. 15, no. 3, pp. 289–308, 2003.
- [12] M. Madkour, D. E. L. Ghanami, A. Maach et al., "Context-aware service adaptation: an approach based on fuzzy sets and service composition," *Journal of Information Science & Engineering*, vol. 29, no. 1, 2013.
- [13] T. Patkos, A. Bikakis, G. Antoniou et al., "Design and challenges of a semantics-based framework for context-aware services," *International Journal of Reasoning-Based Intelligent Systems*, vol. 1, no. 1, pp. 18–30, 2009.
- [14] K. Henricksen and J. Indulska, "Developing context-aware pervasive computing applications: models and approach," *Pervasive and Mobile Computing*, vol. 2, no. 1, pp. 37–64, 2006.
- [15] L. Bálint, "Adaptive human-computer interfaces for man-machine interaction in computer-integrated systems," *Computer Integrated Manufacturing Systems*, vol. 8, no. 2, pp. 133–142, 1995.
- [16] ATTN:ATZL-SWW U.S. Army Combined Arms Center, "Battlefield visualization concept. Department of the Army-TRADOC Pamphlet 525-70," 1995.
- [17] A. Savidis and C. Stephanidis, "Inclusive development: software engineering requirements for universally accessible interactions," *Interacting with Computers*, vol. 18, no. 1, pp. 71–116, 2006.
- [18] A. P. Krkkinen, "Boosting intelligence analysis process and situation awareness using the self-organizing map," in *Defense, Security, and Sensing. International Society for Optics and Photonics*, Proceedings of the SPIE, 2009.

## Research Article

# Mixture Augmented Lagrange Multiplier Method for Tensor Recovery and Its Applications

Huachun Tan,<sup>1</sup> Bin Cheng,<sup>1</sup> Jianshuai Feng,<sup>2</sup> Li Liu,<sup>3</sup> and Wuhong Wang<sup>1</sup>

<sup>1</sup> Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

<sup>2</sup> Integrated Information System Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

<sup>3</sup> Marvell Semiconductor Inc., 5488 Marvell LN, Santa Clara, CA 95054, USA

Correspondence should be addressed to Huachun Tan; tanhc00@gmail.com

Received 30 November 2013; Accepted 30 January 2014; Published 17 March 2014

Academic Editor: Huimin Niu

Copyright © 2014 Huachun Tan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The problem of data recovery in multiway arrays (i.e., tensors) arises in many fields such as computer vision, image processing, and traffic data analysis. In this paper, we propose a scalable and fast algorithm for recovering a low- $n$ -rank tensor with an unknown fraction of its entries being arbitrarily corrupted. In the new algorithm, the tensor recovery problem is formulated as a mixture convex multilinear Robust Principal Component Analysis (RPCA) optimization problem by minimizing a sum of the nuclear norm and the  $\ell_1$ -norm. The problem is well structured in both the objective function and constraints. We apply augmented Lagrange multiplier method which can make use of the good structure for efficiently solving this problem. In the experiments, the algorithm is compared with the state-of-art algorithm both on synthetic data and real data including traffic data, image data, and video data.

## 1. Introduction

A tensor is a multidimensional array. It is the higher-order generalization of vector and matrix, which has many applications in information sciences, computer vision, graph analysis [1], and traffic data analysis [2–4]. In the real world, as the size and the amount of redundancy of the data increase fast and nearly all of the existing high-dimensional real world data either have the natural form of tensor (e.g., multichannel images) or can be grouped into the form of tensor (e.g., tensor face [5], traffic data tensor model [2–4], and videos), challenges come up in many scientific areas when someone confronts with the high-dimensional real world data. Because of some reasons, one wants to capture the underlying low-dimensional structure of the tensor data or seeks to detect the irregular sparse patterns of the tensor data, such as image compression [6], foreground segmentation [7], saliency detection [8], and traffic data completion [2, 3]. As a consequence, it is desirable to develop algorithms that can capture the low-dimensional structure or the irregular sparse patterns in the high-dimensional tensor data.

In the two-dimensional case, that is, the matrix case, the “rank” and “sparsity” are the most useful tools for matrix-valued data analysis. Chandrasekaran et al. [9] proposed the concept of “rank-sparse incoherence” to depict the fundamental identifiability of recovering the low-rank and sparse components. Wright et al. [10] and Candes et al. [11] demonstrated that if the irregular sparse matrix  $S$  is sufficiently sparse (relative to the rank of  $L$ ), one can accomplish the sparse and low-rank recovery by solving the following convex optimization problem:

$$\begin{aligned} \min_{L, S} : & \quad \|L\|_* + \mu \|S\|_1, \\ \text{s.t.} & \quad A = L + S, \end{aligned} \quad (1)$$

where  $A \in \mathbb{R}^{m \times n}$  is the given matrix to be recovered;  $L \in \mathbb{R}^{m \times n}$  is the low-rank component of  $A$ ;  $S \in \mathbb{R}^{m \times n}$  is the sparse component of  $A$ ;  $\|\cdot\|_*$  denotes the nuclear norm defined by the sum of all singular values;  $\|\cdot\|_1$  denotes the sum of the absolute values of matrix entries;  $\mu$  is a positive weighted parameter. This optimization method is called

the Robust Principal Component Analysis [10, 11] (RPCA) or the Principal Component Pursuit (PCP) due to its ability of exactly recovering the underlying low-rank matrix even in the presence of being corrupted by large entries or outliers.

Although the low-rank matrix recovery problem has been well studied, there is not much work on tensors. Li et al. [12] derived a method for the optimal Rank  $-(R_1, R_2, R_3)$  tensor decomposition model. Considering a real  $n$ -mode tensor  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}$ , the best Rank  $-(R_1, R_2, R_3)$  approximation is to find a tensor  $\mathcal{A}^* \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}$  with prespecified  $\text{rank}_k(A_{(k)}) = R_k$  that minimizes the least-squares cost function:

$$\begin{aligned} \min_{\mathcal{A}^*} \quad & f(\mathcal{A}^*) = \|\mathcal{A} - \mathcal{A}^*\|_F^2, \\ \text{s.t.} \quad & \text{Rank}_k(A_k) = R_k, \quad \forall k. \end{aligned} \quad (2)$$

The  $n$ -rank conditions imply that  $\mathcal{A}^*$  should have the Tucker decomposition [13] as  $\mathcal{A}^* \approx \times_1 U_1 \times_2 U_2, \dots, \times_k U_k, \dots, \times_n U_n$ . For the application, they applied the model to the high-dimensional tensor-like visual data by dividing the observed tensor into a low-dimensional structure plus unbounded but sparse irregular patterns:  $\mathcal{A} = \mathcal{L} + \mathcal{S}$ . By assumption that the  $n$ -rank of  $\mathcal{L}$  should be small and the corruption  $\mathcal{S}$  is bounded, the original function is as follows:

$$\begin{aligned} \min_{\mathcal{L}} \quad & \sum_i \lambda_i \text{rank}_i(\mathcal{L}), \\ \text{s.t.} \quad & \|\mathcal{L} - \mathcal{A}\|_F^2 < \varepsilon^2. \end{aligned} \quad (3)$$

In order to solve the problem, they made some conversions to (3) and extended the matrix robust PCA problem to the tensor case. A relaxation technique was used to separate the dependant relationships and the block coordinate descent (BCD) method was used to solve the low- $n$ -rank tensor recovery problem. Then they proposed the rank sparsity tensor decomposition (RSTD) algorithm. In fact, their algorithm can be seen as a basic version of Lagrange multiplier method. Although being simple and provably correct, the RSTD algorithm requires a very large number of iterations to converge and it is difficult to choose the parameters for speedup. Besides, due to the property of the basic Lagrange multiplier method, the accuracy of results needs to be improved.

In this paper, a new algorithm for low- $n$ -tensor recovery, which is termed as Mixture Augmented Lagrange Multiplier Method for Tensor Recovery (MALM-TR), is proposed. In the new algorithm, analogy to the RSTD [12], we convert the tensor recovery problem into a mixture convex optimization problem by adopting the relaxation technique strategy which eliminates the interdependent trace norm and  $\ell_1$  norm constrain. Actually, the elements involved in problem are all in matrix case. Thus, it can be treated as a multilinear extension of the RPCA problem and subsumes the matrix RPCA problem as a special case. Lin et al. [14] have proved that the matrix RPCA problem can be solved by ALM with achieving higher precision, being less storage/memory demanding, and having a pleasing Q-linear convergence

speed. Inspired by these merits of ALM, we try to extend the augmented Lagrange multiplier method (ALM) to the multilinear RPCA problem and prove that ALM is not only fit to solve the matrix RPCA problem but also suitable to solve the multilinear RPCA problem.

For the usage of this algorithm, it is applied to real world data recovery including traffic data recovery, image restoration, and background modeling.

In traffic data analysis area, due to detector and communication malfunctions, traffic data often confronts with the noising data phenomenon, especially the outlier value noise, which has a great impact on the performance of Intelligent Transportation System (ITS). Therefore, it is essential to solve the issues caused by outlier data in order to fully explore the applicability of the data and realize the ITS applications. In the application part of this paper, we introduce the tensor form to model the traffic data, which can encode the multimode (e.g., week, day, record) correlations of the traffic data simultaneously and preserve the multiway nature of traffic data. For example, it is assumed that a loop detector collects traffic volume data every 15 minutes. Thus, it will have 96 records in a day. If we have 20 weeks traffic volume data, these data can be formed into a tensor of size  $20 \times 7 \times 96$ . Then, the proposed tensor-based method which can well mine the multimode correlations of traffic data mentioned above is used to remove outlier noise of the traffic data.

It is observed that the multichannel image can be seen as a tensor with multidimensions. For example, RGB image has three channels including Red channel, Green channel, and Black channel. Thus, it can be represented as width  $\times$  height  $\times$  3 which is a 3-dimensional tensor. For the application, the proposed method is used to remove the noise of the image. Though the method would not be reasonable for some natural images, it has many applications for visual data such as structured images (e.g., the façade image), CT/MRI data, and multispectral image. Besides images, video data can be grouped into the form of tensor. For example, there is a video which has 300 gray frames and each of which is in size of  $200 \times 200$ . These video data can form a tensor of size  $200 \times 200 \times 300$ . For the video application, the proposed method will be used for background modeling.

The rest of the paper is organized as follows. Section 2 presents some notations and states some basic properties of tensors. Section 3 discusses the detailed process of our proposed algorithm. Section 4 tests the algorithm on different settings, varying from simulated data to applications in computer vision, image processing, and traffic recovery. Finally, some concluding remarks are provided in Section 5.

## 2. Notation and Basics on Tensor Model

In this paper, the nomenclatures and the notations in [1, 12] on tensors are partially adopted. Scalars are denoted by lowercase letters ( $a, b, c, \dots$ ), vectors by bold lowercase letters ( $\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$ ), and matrices by uppercase letters ( $A, B, C, \dots$ ). Tensors are written as calligraphic letters ( $\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$ ).  $N$ -mode tensors are denoted as  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ . The elements of an  $N$ -mode are denoted as  $a_{i_1 \dots i_k \dots i_n}$ , where  $1 \leq i_k \leq I_k$ ,

$1 \leq K \leq N$ . The mode- $n$  unfolding (also called matricization or flattening) of a tensor  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  is defined as  $\text{unfold}(\mathcal{A}, n) = A_{(n)}$ . The tensor element  $(i_1, i_2, \dots, i_N)$  is mapped to the matrix element  $(i_n, j)$ , where

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^N (i_k - 1) J_k, \quad \text{with } J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m. \quad (4)$$

Therefore,  $A_{(n)} \in \mathbb{R}^{I_n \times J}$ , where  $J = \prod_{\substack{k=1 \\ k \neq n}}^N I_k$ . Accordingly, its inverse operator fold can be defined as  $\text{fold}(A_{(n)}, n) = \mathcal{A}$ .

The  $n$ -rank of a  $N$ -dimensional tensor  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ , denoted by  $r_n$ , is the rank of the mode- $n$  unfolding matrix  $A_{(n)}$ :

$$r_n = \text{rank}_n(\mathcal{A}) = \text{rank}(A_{(n)}). \quad (5)$$

If the  $n$ -rank is very small related to the size of the tensor, we call it low- $n$ -rank tensor.

The inner product of two same-size tensors  $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  is defined as the sum of the products of their entries, that is,

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} a_{i_1 \dots i_N} b_{i_1 \dots i_N}. \quad (6)$$

The corresponding Frobenius norm is  $\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$ . Besides, the  $\ell_0$  norm of a tensor  $\mathcal{A}$ , denoted by  $\|\mathcal{A}\|_0$ , is the number of nonzero elements in  $\mathcal{A}$  and the  $\ell_1$  norm is defined as  $\|\mathcal{A}\|_1 := \sum_{i_1 \dots i_N} |a_{i_1 \dots i_N}|$ . It is clear that  $\|\mathcal{A}\|_F = \|A_{(n)}\|_F$ ,  $\|\mathcal{A}\|_0 = \|A_{(n)}\|_0$  and  $\|\mathcal{A}\|_1 = \|A_{(n)}\|_1$  for any  $1 \leq n \leq N$ .

The  $n$ -mode (matrix) product of a tensor  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  with a matrix  $M \in \mathbb{R}^{J \times I_n}$  is denoted by  $\mathcal{A} \times_n M$  and is of size  $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$ . In terms of flattened matrix, the  $n$ -mode product can be expressed as

$$\mathcal{Y} = \mathcal{A} \times_n M \iff Y_{(n)} = M A_{(n)}. \quad (7)$$

### 3. MALM-TR

This section is separated into 2 parts. In Section 3.1, we convert the low- $n$ -tensor recovery problem into a multilinear RPCA problem. Section 3.2 simply introduces the ALM approach, extends ALM approach to solve the multilinear RPCA problem, and presents the details of the proposed algorithm.

**3.1. The Multilinear RPCA Problem.** The derivation starts with the general version [10] of matrix recovery problem:

$$\begin{aligned} \min_{L, S} : \quad & \text{rank}(L) + \lambda \|S\|_0, \\ \text{s.t.} \quad & L + S = A, \end{aligned} \quad (8)$$

where  $A \in \mathbb{R}^{m \times n}$  is the given matrix to be recovered;  $L \in \mathbb{R}^{m \times n}$  is the low-rank component of  $A$ ;  $S \in \mathbb{R}^{m \times n}$  is the sparse component of  $A$ ;  $\text{rank}(L)$  denotes the rank of  $L$ ;  $\|\cdot\|_0$  denotes the number of nonzero matrix entries;  $\lambda$  is a positive weighted

parameter. The higher-order tensor recovery problem can be generated from the matrix (i.e., 2nd-order tensor) case by utilizing the form of (8), leading to the formulation of the following:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S}} \quad & \text{rank}_{\text{CP}}(\mathcal{L}) + \lambda \|\mathcal{S}\|_0, \\ \text{s.t.} \quad & \mathcal{L} + \mathcal{S} = \mathcal{A}, \end{aligned} \quad (9)$$

where  $\mathcal{L}, \mathcal{S}, \mathcal{A}$  are  $n$ -mode tensors with identical size in each mode.  $\mathcal{A}$  is the observed tensor data.  $\mathcal{L}$  and  $\mathcal{S}$  represent the correspondent structured part and irregular sparse part, respectively.  $\text{rank}_{\text{CP}}(\mathcal{L})$  is the minimum number of rank-1 tensors that generates  $\mathcal{L}$  as their sum [15, 16]. However, (9) is unsolvable because there is no straightforward algorithm to determine the CP-rank of a specific given tensor and the  $\ell_0$  norm is highly nonconvex. But when the given tensor is a low- $n$ -rank tensor we can use the  $n$ -rank of unfolding of a tensor  $\mathcal{A}$  instead of CP-rank of tensor to capture the global information of the given tensor. Therefore, we can minimize the  $n$ -ranks of the given tensor, respectively, instead of minimizing the CP-rank to solve the tensor completion problem. Obviously,  $\|\mathcal{S}\|_0$  is equal to  $\|S_{(n)}\|_0$ . As a result, a function  $F$  which minimizes all the  $n$ -ranks of the given tensor to replace (9) is obtained as follows:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}} \quad & \left\{ F(\text{rank}(L_{(1)}), \text{rank}(L_{(2)}), \right. \\ & \left. \dots, \text{rank}(L_{(N)})) + \frac{1}{n} \sum_{i=1}^n \eta_i \|S_{(i)}\|_0 \right\}, \quad (10) \\ \text{s.t.} \quad & L_{(i)} + S_{(i)} = A_{(i)}, \end{aligned}$$

where  $L_{(i)}, S_{(i)}$  are the mode- $i$  unfoldings of  $\mathcal{L}$  and  $\mathcal{S}$ . Equation (10) is a highly nonconvex optimization problem, and no efficient solution is known due to the nonconvexness of the matrix rank and  $\ell_0$  norm. Fortunately, it is a fact that the nuclear norm and  $\ell_1$  norm are the tightest convex approximation of rank and  $\ell_0$  norm [10, 11], respectively. By replacing rank with nuclear norm and replacing  $\ell_0$ -norm with  $\ell_1$ -norm, a tractable convex optimization problem can be obtained:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}} \quad & \left\{ F(\|L_{(1)}\|_*, \|L_{(2)}\|_*, \dots, \|L_{(n)}\|_*) \right. \\ & \left. + \frac{1}{n} \sum_{i=1}^n \eta_i \|S_{(i)}\|_1 \right\}, \quad (11) \\ \text{s.t.} \quad & L_{(i)} + S_{(i)} = A_{(i)}. \end{aligned}$$

In order to utilize the information of each mode as much as possible, the  $n$ -rank minimization problems of each mode are combined by weighted parameters to replace the function  $F$

with  $(1/n) \sum_i^n \lambda_i (\|L_{(i)}\|_*)$  which is defined in [17, 18]. Thus, the tensor completion problem becomes

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S}} \quad & \frac{1}{n} \sum_{i=1}^n \lambda_i \|L_{(i)}\|_* + \frac{1}{n} \sum_{i=1}^n \eta_i \|S_{(i)}\|_1, \\ \text{s.t.} \quad & L_{(i)} + S_{(i)} = A_{(i)}. \end{aligned} \quad (12)$$

Problem (12) is still hard to solve due to the inter-dependent trace norm and  $\ell_1$  norm constraint. In order to simplify the problem, we introduce additional auxiliary matrix  $M_{(i)} = L_{(i)}$ ,  $N_{(i)} = S_{(i)}$ . Then, we relax the equality constrains by  $\|L_{(i)} - M_{(i)}\|_F^2 \leq \varepsilon_2$  and  $\|S_{(i)} - N_{(i)}\|_F^2 \leq \varepsilon_3$ . It is easy to check that  $\|A_{(i)} - L_{(i)} - S_{(i)}\|_F^2 \leq \varepsilon_1$  corresponds to the stable Principle Component Pursuit (sPCP) in the matrix case [19]. Finally, we get the relaxed form of (12) which can be seen as a multilinear RPCA problem:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S}} \quad & \frac{1}{n} \sum_{i=1}^n \lambda_i \|M_{(i)}\|_* + \frac{1}{n} \sum_{i=1}^n \eta_i \|N_{(i)}\|_1, \\ \text{s.t.} \quad & \|A_{(i)} - L_{(i)} - S_{(i)}\|_F^2 \leq \varepsilon_1, \\ & \|L_{(i)} - M_{(i)}\|_F^2 \leq \varepsilon_2, \\ & \|S_{(i)} - N_{(i)}\|_F^2 \leq \varepsilon_3. \end{aligned} \quad (13)$$

**3.2. Optimization Process.** In [20], the general method of augmented Lagrange multipliers is introduced for solving constrained optimization problems of the kind:

$$\begin{aligned} \min \quad & f(X), \\ \text{s.t.} \quad & h(X) = 0, \end{aligned} \quad (14)$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  and  $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the augmented Lagrange function is defined as

$$(X, Y, \mu) = f(X) + \langle Y, h(X) \rangle + \frac{\omega}{2} \|h(X)\|_F^2, \quad (15)$$

where  $\omega$  is a positive scalar, and then the optimization problem can be solved via the method of augmented Lagrange multipliers (see [21] for more details).

It is observed that (13) is well structured and the separable structure emerges in both the objective function and constraint conditions. We convert (9) into its augmented

Lagrange form with proper  $\alpha_i$ ,  $\beta_i$ , and  $\gamma_i$ . The augmented Lagrange of (13) is

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S}, M_{(i)}, N_{(i)}} \quad & \sum_{i=1}^n \lambda_i \|M_{(i)}\|_* + \sum_{i=1}^n \eta_i \|N_{(i)}\|_1 \\ & + \sum_{i=1}^n \langle Y_i, L_{(i)} - M_{(i)} \rangle + \sum_{i=1}^n \frac{\alpha_i}{2} \|L_{(i)} - M_{(i)}\|_F^2 \\ & + \sum_{i=1}^n \langle Z_i, S_{(i)} - N_{(i)} \rangle + \sum_{i=1}^n \frac{\beta_i}{2} \|S_{(i)} - N_{(i)}\|_F^2 \\ & + \sum_{i=1}^n \langle W_i, A_{(i)} - M_{(i)} - N_{(i)} \rangle \\ & + \sum_{i=1}^n \frac{\gamma_i}{2} \|A_{(i)} - M_{(i)} - N_{(i)}\|_F^2. \end{aligned} \quad (16)$$

Equation (16) can be simplified into its equivalent form:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S}, M_{(i)}, N_{(i)}} \quad & \sum_{i=1}^n \lambda_i \|M_{(i)}\|_* + \sum_{i=1}^n \eta_i \|N_{(i)}\|_1 \\ & + \sum_{i=1}^n \frac{\alpha_i}{2} \left\| M_{(i)} - L_{(i)} - \frac{Y_i}{\alpha_i} \right\|_F^2 \\ & + \sum_{i=1}^n \frac{\beta_i}{2} \left\| N_{(i)} - S_{(i)} - \frac{Z_i}{\beta_i} \right\|_F^2 \\ & + \sum_{i=1}^n \frac{\gamma_i}{2} \left\| M_{(i)} + N_{(i)} - A_{(i)} - \frac{W_i}{\gamma_i} \right\|_F^2. \end{aligned} \quad (17)$$

The core idea of solving the optimization problem in (17) is to optimize a group of variables while fixing the other groups. The variables in the optimization are  $M_{(1)}, \dots, M_{(n)}, N_{(1)}, \dots, N_{(n)}, S_{(i)}, L_{(i)}$  which can be divided into  $2n + 2$  groups. To achieve the optimal solution, the method estimates  $M_{(i)}, N_{(i)}, S_{(i)}, L_{(i)}$  sequentially, followed by certain refinement in each iteration.

**Computing  $M_{(i)}$ .** The optimal  $M_{(i)}$  with all other variables fixed is the solution to the following subproblem:

$$\begin{aligned} \min_{M_{(i)}} \quad & \lambda_i \|M_{(i)}\|_* + \frac{\alpha_i}{2} \left\| M_{(i)} - L_{(i)} - \frac{Y_i}{\alpha_i} \right\|_F^2 \\ & + \frac{\gamma_i}{2} \left\| M_{(i)} + N_{(i)} - A_{(i)} - \frac{W_i}{\gamma_i} \right\|_F^2. \end{aligned} \quad (18)$$

As shown in [22], the optimal solution of (18) is given by

$$M_{(i)} = U_i D_{\lambda_i / \alpha_i + \gamma_i}(\Lambda) V_i^T, \quad (19)$$

where  $U_i \Lambda V_i^T$  is the singular value decomposition given by

$$U_i \Lambda V_i^T = \frac{(\alpha_i (L_{(i)} + Y_i / \alpha_i) + \gamma_i (A_{(i)} - N_{(i)} + W_i / \gamma_i))}{(\alpha_i + \gamma_i)}, \quad (20)$$

**Input:**  $n$ -mode tensor  $\mathcal{A}$ .

**Parameters:**  $\alpha, \beta, \gamma, \lambda, \eta, \rho$ .

(1) Initialization:  $M_{(i)} = L_{(i)}, N_{(i)} = 0, k = 1, \rho > 0$ .

(2) Repeat until convergence

(3) for  $i = 1$  to  $n$

(4)  $M_{(i)}^{k+1} = U_i D_{\lambda_i/\alpha_i + \gamma_i}(\Lambda) V_i^T$ ,

Where  $U_i \Lambda V_i^T = (\alpha_i (L_{(i)}^k + Y_i^k / \alpha_i) + \gamma_i (A_{(i)}^k - N_{(i)}^k + W_i^k / \gamma_i)) / (\alpha_i + \gamma_i)$ .

(5)  $N_{(i)}^{k+1} = \mathcal{D}_{\eta_i/\beta_i + \gamma_i} \left( (\beta_i (S_{(i)}^k + Z_i^k / \beta_i) + \gamma_i (A_{(i)}^k - M_{(i)}^k + W_i^k / \gamma_i)) / (\beta_i + \gamma_i) \right)$ .

(6)  $Y_i^{k+1} = Y_i^k + \alpha_i (L_{(i)}^{k+1} - M_{(i)}^{k+1})$ ,

(7)  $Z_i^{k+1} = Z_i^k + \beta_i (S_{(i)}^{k+1} - N_{(i)}^{k+1})$ ,

(8)  $W_i^{k+1} = W_i^k + \gamma_i (A_{(i)} - M_{(i)}^{k+1} - N_{(i)}^{k+1})$ .

(9) end for

(10)  $\mathcal{L}^{k+1} = \sum_{i=1}^n \alpha_i \text{fold} (M_{(i)}^{k+1} - Y_i^{k+1} / \alpha_i, i) / \sum_{i=1}^n \alpha_i$ ,

(11)  $\mathcal{S}^{k+1} = \sum_{i=1}^n \beta_i \text{fold} (N_{(i)}^{k+1} - Z_i^{k+1} / \beta_i, i) / \sum_{i=1}^n \beta_i$ .

(12)  $\alpha = \rho \alpha, \beta = \rho \beta, \gamma = \rho \gamma$ .

(13)  $k = k + 1$ .

(14) End

**Output:**  $n$ -mode tensor  $\mathcal{L}, \mathcal{S}$ .

ALGORITHM 1: MALM-TR: MALM for tensor recovery.

and  $\mathcal{D}_\tau$  is the ‘‘shrinkage’’ operation. The ‘‘shrinkage’’ operator  $\mathcal{D}_\tau(x)$  with  $\tau > 0$  is defined as

$$\mathcal{D}_\tau(x) = \begin{cases} x - \tau, & \text{if } x > \tau \\ x + \tau, & \text{if } x < -\tau \\ 0, & \text{otherwise.} \end{cases} \quad (21)$$

The operator can be extended to the matrix or tensor case by performing the shrinkage operator towards each element.

*Computing  $N_{(i)}$ .* The optimal  $N_{(i)}$  with all other variables fixed is the solution to the following subproblem:

$$\begin{aligned} \min_{N_{(i)}} : & \eta_i \|N_{(i)}\|_1 + \frac{\beta_i}{2} \left\| N_{(i)} - S_{(i)} - \frac{Z_i}{\beta_i} \right\|_F^2 \\ & + \frac{\gamma_i}{2} \left\| M_{(i)} + N_{(i)} - A_{(i)} - \frac{W_i}{\gamma_i} \right\|_F^2. \end{aligned} \quad (22)$$

By the well-known  $\ell_1$  norm minimization [23], the optimal solution of (22) is

$$\begin{aligned} N_{(i)} = & \mathcal{D}_{\eta_i/\beta_i + \gamma_i} \\ & \times \left( \frac{(\beta_i (S_{(i)} + Z_i/\beta_i) + \gamma_i (A_{(i)} - M_{(i)} + W_i/\gamma_i))}{(\beta_i + \gamma_i)} \right). \end{aligned} \quad (23)$$

*Computing  $\mathcal{S}_{(i)}$ .* The optimal  $\mathcal{S}$  with all other variables fixed is the solution to the following subproblem:

$$\min_{\mathcal{S}} : \sum_{i=1}^n \frac{\beta_i}{2} \left\| N_{(i)} - S_{(i)} - \frac{Z_i}{\beta_i} \right\|_F^2. \quad (24)$$

It is easy to show that the solution to (24) is given by

$$\mathcal{S} = \sum_{i=1}^n \frac{\beta_i \text{fold} (N_{(i)} - Z_i/\beta_i, i)}{\sum_{i=1}^n \beta_i}. \quad (25)$$

*Computing  $L_{(i)}$ .* The optimal  $\mathcal{L}$  with all other variables fixed is the solution to the following subproblem:

$$\min_{\mathcal{L}} : \sum_{i=1}^n \frac{\alpha_i}{2} \left\| M_{(i)} - L_{(i)} - \frac{Y_i}{\alpha_i} \right\|_F^2. \quad (26)$$

It is easy to show that the solution to (26) is given by

$$\mathcal{L} = \sum_{i=1}^n \frac{\alpha_i \text{fold} (M_{(i)} - Y_i/\alpha_i, i)}{\sum_{i=1}^n \alpha_i}. \quad (27)$$

The pseudo-code of the proposed MALM-TR algorithm is summarized in Algorithm 1.

Under some rather general conditions, when  $\{\alpha, \beta, \gamma\}$  is an increasing sequence and both the objective function and the constraints are continuously differentiable functions, it has been proven in [20] that the Lagrange multipliers  $\{Y_i, Z_i, W_i\}$  produced by Algorithm 1 converge Q-linearly to the optimal solution when  $\{\alpha, \beta, \gamma\}$  is bounded and super-Q-linearly when  $\{\alpha, \beta, \gamma\}$  is unbounded. Another merit of MALM-TR is that the optimal step size to update  $\{Y_i, Z_i, W_i\}$  is proven to be the chosen penalty parameters  $\{\alpha, \beta, \gamma\}$ , making the parameter tuning much easier. A third merit of MALM-TR is that the algorithm converges to the exact optimal solution, even without requiring  $\{\alpha, \beta, \gamma\}$  to approach infinity [20].

## 4. Experiments

In this section, using both the numerical simulations and the real world data, we evaluate the performance of our proposed

TABLE 1:  $\mathcal{A}_0 \in \mathbb{R}^{40 \times 40 \times 40}$ ,  $n$ -rank = [5, 5, 5], spr = 5%, 15%, 25%, 35%.

spr	Algorithm: MALM-TR				Algorithm: RSTD			
	RSE- $\mathcal{L}_0$ ( $e-5$ )	RSE- $\mathcal{S}_0$ ( $e-5$ )	# iter	Time (s)	RSE- $\mathcal{L}_0$ ( $e-5$ )	RSE- $\mathcal{S}_0$ ( $e-5$ )	# iter	Time (s)
0.05	0.01	0.006	136	17.7	450	430	226	19.5
0.15	0.06	0.02	167	19.5	910	440	330	27.4
0.25	1.1	0.3	281	29.1	1510	490	714	44.4
0.35	2010	390	450	41.9	5620	1140	608	44.1

TABLE 2:  $\mathcal{A}_0 \in \mathbb{R}^{40 \times 40 \times 40}$ ,  $n$ -rank = [10, 10, 10], spr = 5%, 10%, 15%, 20%.

spr	Algorithm: MALM-TR				Algorithm: RSTD			
	RSE- $\mathcal{L}_0$ ( $e-5$ )	RSE- $\mathcal{S}_0$ ( $e-5$ )	# iter	Time (s)	RSE- $\mathcal{L}_0$ ( $e-5$ )	RSE- $\mathcal{S}_0$ ( $e-5$ )	# iter	Time (s)
0.05	0.2	0.1	243	29.4	400	230	411	33.8
0.10	10.1	4.2	323	43.5	520	270	568	51.3
0.15	120	42	596	62.5	650	260	1103	68.5
0.20	1250	370	972	88.3	2650	780	1235	88.1

algorithm and then compare the results with RSTD on the low- $n$ -rank tensor recovery problem.

In all the experiments, the Lanczos bidiagonalization algorithm with partial reorthogonalization [24] is adopted to obtain a few singular values and vectors in all iterations. A major challenge of our algorithm is the selection of parameters. We simply set the parameters  $\alpha = \beta = \gamma = [I_1/I_{\max}, I_2/I_{\max}, \dots, I_n/I_{\max}]^T$  for all experiments, where  $I_{\max} = \max\{I_i\}$ . Similarly, we choose  $\eta = 1/\sqrt{I_{\max}}$  as suggested in [11] and tune  $\lambda$  with the change of  $\eta$ . For comparing with RSTD [12], we also use the difference of  $\mathcal{L}$  and  $\mathcal{S}$  in successive iterations against a certain tolerance as the stopping criterion. All the experiments are conducted and timed on the same desktop with a Pentium (R) Dual-Core 2.50 GHz CPU that has 4 GB memory, running on Windows 7 and Matlab.

**4.1. Numerical Simulations.** A low- $n$ -rank tensor  $\mathcal{L}_0$  is generated as follows. The  $N$ -way Tensor Toolbox [25] is used to generate a third-order tensor with the size of  $I_1 \times I_2 \times I_3$  and the relative small  $n$ -rank  $[r_1 r_2 r_3]$ . The generated tensor is in Tucker model [13] described as  $\mathcal{L}_0 = \mathcal{C}_{\times 1} X_{\times 2} Y_{\times 3} Z$ . To impose these rank conditions,  $\mathcal{C}$  is  $\mathbb{R}^{r_1 \times r_2 \times r_3}$  core tensor with each entry being sampled independently from a standard Gaussian distribution  $\mathcal{N}(0, 1)$ ,  $X, Y$  and  $Z$  are  $I_1 \times r_1, I_2 \times r_2, I_3 \times r_3$  factor matrices generated by randomly choosing each entry from  $\mathcal{N}(0, 1)$ . Here without loss of generality we make the factor matrices orthogonal. But one major difference is that the  $n$ -ranks are always different along each mode while the column rank and row rank of a matrix are equal to each other. For simplicity, in this paper we set the mode- $n$  ranks with the same value.

The entries of sparse tensor  $\mathcal{S}_0$  are independently distributed, each taking on value 0 with probability  $1 - \text{spr}$ , and each taking on impulsive value with probability  $\text{spr}$ . The recovered tensor  $\mathcal{A}_0$  is generated as  $\mathcal{A}_0 = \mathcal{L}_0 + \mathcal{S}_0$ .

The simulated tensor used in the experiments is of size  $40 \times 40 \times 40$ , varying the  $n$ -rank  $r$  and the sparse ratio  $\text{spr}$ . The parameters are adjusted according to the different  $r$  and  $\text{spr}$ . The quality of recovery is measured by the relative square error (RSE) to  $\mathcal{L}_0$  and  $\mathcal{S}_0$ , which is defined as

$$\text{RSE-}\mathcal{L}_0 = \frac{\|\widehat{\mathcal{L}} - \mathcal{L}_0\|_F}{\|\mathcal{L}_0\|_F}, \quad \text{RSE-}\mathcal{S}_0 = \frac{\|\widehat{\mathcal{S}} - \mathcal{S}_0\|_F}{\|\mathcal{S}_0\|_F}. \quad (28)$$

Tables 1 and 2 present the average results (across 10 instances) for different sparse ratio. The results demonstrate that our proposed algorithm MALM-TR outperforms RSTD on either efficiency or accuracy.

**4.2. Image Restoration.** One straightforward application of our algorithm is the image restoration. Same as [12] pointed, our algorithm also assumes the image to be well structured. Though the assumption would not be reasonable for some natural images, it has many applications for visual data such as structured images (e.g., the façade image), CT/MRI data, and multispectral image. In experiments, we apply the algorithm on image restoration of the façade image, which is also used in [12, 17]. We add different percent of random impulsive noise to the image and compare MALM-TR with RSTD. The results produced by both algorithms are shown in Figure 1.

**4.3. Background Modeling.** Another application of our algorithm is to estimate a good model for the background variations in a scene (i.e., background modeling). In this situation, it is natural to model the background variation as approximately low rank. Foreground objects generally occupy only a fraction of the image pixels and hence can be treated as sparse part.

We test our algorithm using an example from [26] and compare with RSTD [12]. The visual comparisons of the background modeling are shown in Figure 2. It is observed that

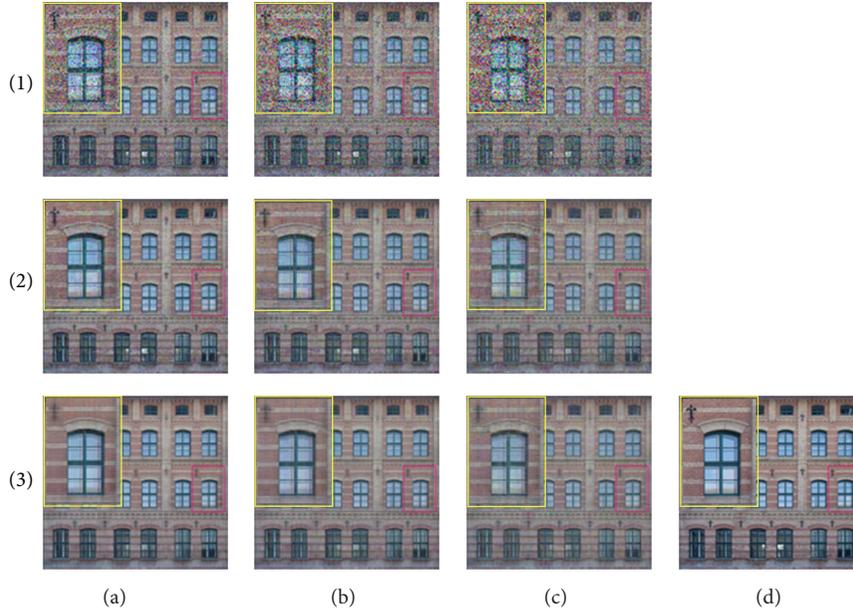


FIGURE 1: Comparisons in terms of visual effects. The rows (1), (2), and (3) correspond to the images before recovery, the obtained results by MALM-TR and RSTD [12], respectively. The columns (a), (b), and (c) correspond to the images corrupted by 15%, 25%, and 35% sparse impulsive noise, respectively. (d) is the original image.

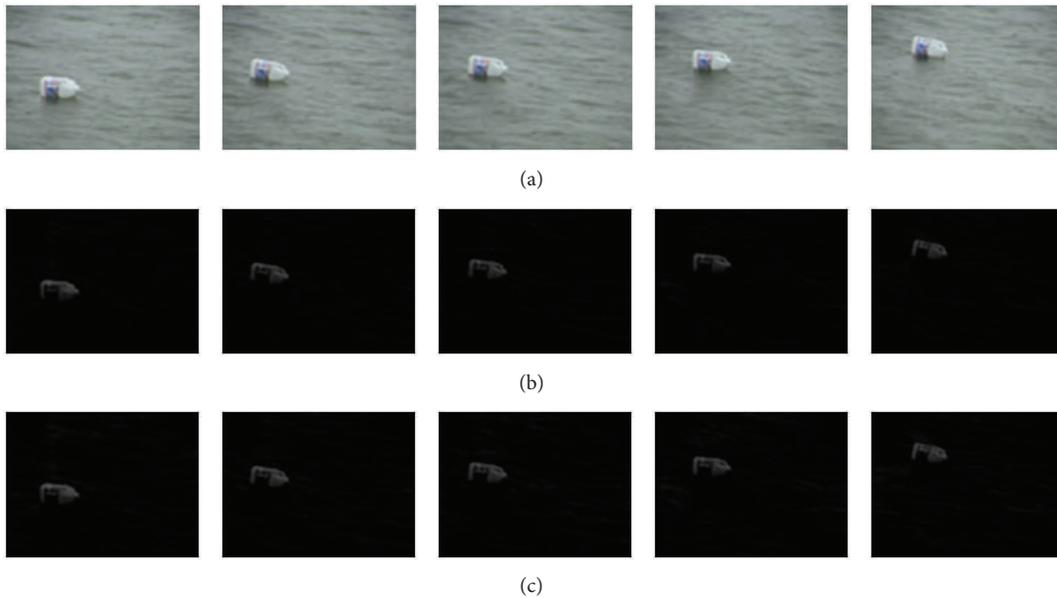


FIGURE 2: Background modeling. Top: original video sequence of a scene. Middle: foreground object recovered by MALM-TR. Bottom: foreground object recovered by RSTD [12]. The results are highlighted for both RSTD and MALM-TR.

our algorithm is effective in separating the background which is even a dynamic scene. The results are also comparable to RSTD.

*4.4. Traffic Data Recovery.* In our previous work [3, 4], we have proposed two tensor-based methods on traffic data application. In [3], a tensor imputation method based on Tucker decomposition is developed to estimate the missing

value. As a fact that the exact coordinate and the number of the missing data in the tensor form can be observed and obtained because if an element in the tensor form is missing, it doesn't have value so we can recognize it easily. While this paper recovers a low- $n$ -rank tensor that is arbitrarily corrupted by a fraction of noise based on the trace norm and  $\ell_1$ -norm optimization. The number and the coordinate of the corrupted data are unknown or not easy to obtain. That

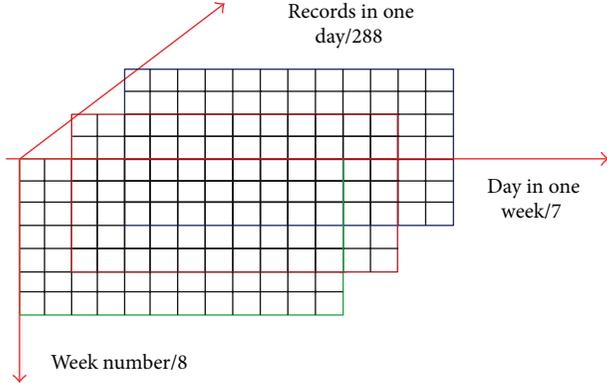
FIGURE 3: Tensor model of size  $8 \times 7 \times 288$ .

TABLE 3: Comparison of RSE on traffic data.

SPR	MALM-TR(RSE)	RSTD(RSE)	Unrecovered(RSE)
0.05	0.0406	0.0525	0.2614
0.10	0.0432	0.0543	0.4117
0.15	0.0593	0.1370	0.5427
0.20	0.0784	0.2198	0.6859
0.25	0.1478	0.3959	0.8218

means that it is hard to recognize the corrupted data, because the corrupted data have values and hardly be separated from the correct data. The problems solved by the two papers are two different problems. Paper [4] is written from the point of traffic data recovery application which is the same problem that will be solved in this section. The main difference of the two proposed methods is how to use the constraint condition  $\mathcal{A} = \mathcal{L} + \mathcal{S}$ . Reference [4] puts the constraint condition to the minimized function with only one parameter  $\gamma$ , which leads the objective function to contain not only tensor but also matrix. However, as the size and structure of each mode of the given tensor data are not always the same, the contribution of each mode of the tensor to the final result may be different. In order to utilize the information of the constraint condition as much as possible, this paper unfolds constraint condition along each mode and use weighted parameters  $\gamma_i$  to obtain the new constraint condition in matrix versions  $\sum_{i=1}^n (\gamma_i/2) \|A_{(i)} - M_{(i)} - N_{(i)}\|_F^2$  which is put into the minimized function using the augmented Lagrange multiplier strategy. With different objective functions, the optimized process is different too. More details can be found in [4].

In the fourth part of the experiment section, we will apply the proposed algorithm to traffic data recovery. The data used in the experiment are collected by a fixed loop in Sacramento County and downloaded from <http://pems.dot.ca.gov/>. The period of the data lasts for 77 days from March 14 to May 29, 2011. The traffic volume data are recorded every 5 minutes. Therefore, a daily traffic volume series for a loop detector contains 288 records. To finish traffic data recovery by the proposed algorithm, the first step is to convert the mass traffic data into a tensor form. In this part, we choose 8-week

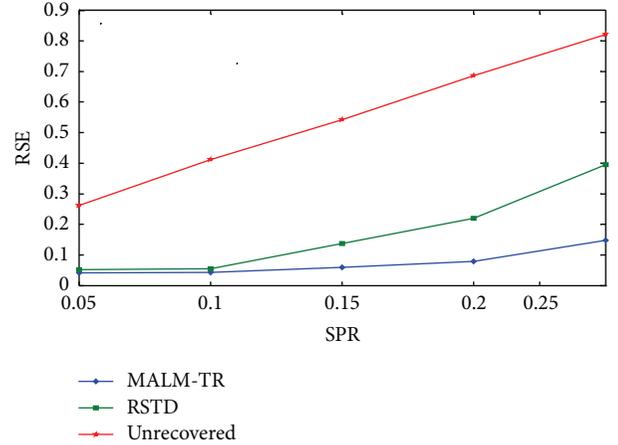


FIGURE 4: Comparison of RSE curves on traffic data.

complete traffic volume data from the 77 days. Then, the 8-week data are formed into a tensor model of size  $8 \times 7 \times 288$  as Figure 3 shows. In this model, “8” stands for 8 weeks, “7” stands for seven days in one week, and “288” stands for 288 records in one day.

In our previous work [3], the similarity coefficient [27] had been used to analyze the high multimode correlation (“link” mode, “week” mode, “day mode,” and “hour” mode) of traffic data from the point of view of statistic characteristic. For the high multicorrelations of traffic data, the tensor form of size  $8 \times 7 \times 288$  can be approximated by a low- $n$ -rank tensor.

According to the above description, the traffic data are reasonably converted into a tensor form which can be approximated by a low- $n$ -rank tensor. In the traffic data recovery experiment, it is assumed that a subset of entries of the traffic data tensor form is corrupted by impulsive noise at random. The ratios of noisy are set from 5% to 25% with the tolerance 5%. Then we compare the proposed method with RSTD algorithm using RSE as the criterion. The criterion is defined as the following function shows:

$$\text{RSE} = \frac{\|\text{recovered data} - \text{original data}\|_F}{\|\text{original data}\|_F}. \quad (29)$$

Table 3 tabulates the RSEs by sparse impulsive noise with different ratio on traffic data. Especially, the unrecovered column presents the RSE between corrupted data and the original data. From data in the table, it is observed that the RSEs obtained by MALM-TR and RSTD are much smaller than the unrecovered data, which means that both algorithm can improve the quality of the corrupted data. Moreover, the RSEs of MALM-TR are smaller than RSTD. From the curves of Figure 4, it is vividly shown that our method performs better than RSTD.

## 5. Conclusion

In this paper, we extend the matrix recovery problem to low- $n$ -rank tensor recovery and propose an efficient algorithm

based on mixture augmented Lagrange multiplier method. The proposed algorithm can automatically separate the low- $n$ -rank tensor data and sparse part. Experiments show that the proposed algorithm is more stable and accurate in most cases and has excellent convergence rate. Different application examples show the broad applicability of our proposed algorithm in computer vision, image processing, and traffic data recovery.

In the future, we would like to investigate how to automatically choose the parameters in our algorithm and develop more efficient method for tensor recovery problem. Also we will explore more applications of our method.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The research was supported by NSFC (Grant nos. 61271376, 51308115, and 91120010), the National Basic Research Program of China (973 Program: no. 2012CB725405), and Beijing Natural Science Foundation (4122067). The authors would like to thank Professor Bin Ran from the University of Wisconsin-Madison and Yong Li from the University of Notre Dame for the suggestive discussions.

## References

- [1] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [2] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41–56, 2011.
- [3] H. Tan, G. Feng, J. Feng, W. Wang, Y. J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C*, vol. 28, pp. 15–27, 2013.
- [4] H. Tan, G. Feng, J. Feng, W. Wang, and Y. J. Zhang, "Traffic volume data outlier recovery via tensor model," *Mathematical Problems in Engineering*, vol. 2013, Article ID 164810, 8 pages, 2013.
- [5] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Proceedings of the European Conference on Computer Vision (ECCV '02)*, pp. 447–460, 2002.
- [6] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Transactions of Image Processing*, vol. 1, no. 2, pp. 244–250, 1992.
- [7] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1778–1792, 2005.
- [8] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [9] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," <http://arxiv.org/abs/0906.2220>.
- [10] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao, "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," in *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS '09)*, pp. 2080–2088, December 2009.
- [11] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" 2009, arXiv:0912.3599v1.
- [12] Y. Li, J. Yan, Y. Zhou, and J. Yang, "Optimum subspace learning and error correction for tensors," in *Proceedings of the 11th European Conference on Computer Vision (ECCV '10)*, Crete, Greece, 2010.
- [13] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [14] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of a corrupted low-rank matrices," *Mathematical Programming*, 2009.
- [15] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.
- [16] R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.
- [17] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," in *Proceedings of the International Conference on Computer Vision (ICCV '09)*, 2009.
- [18] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.
- [19] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, "Stable principal component pursuit," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '10)*, pp. 1518–1522, June 2010.
- [20] D. P. Bertsekas and A. E. Ozdaglar, "Pseudonormality and a lagrange multiplier theory for constrained optimization," *Journal of Optimization Theory and Applications*, vol. 114, no. 2, pp. 287–343, 2002.
- [21] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.
- [22] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [23] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for  $\ell_1$ -minimization: methodology and convergence," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1107–1130, 2008.
- [24] H. D. Simon, "The lanczos algorithm with partial reorthogonalization," *Mathematics of Computation*, vol. 42, pp. 115–142, 1984.
- [25] C. A. Andersson and R. Bro, "The N-way toolbox for MATLAB, chemometrics and intelligent laboratory systems," *Chemometrics & Intelligent Laboratory Systems*, vol. 52, no. 1, pp. 1–25.4, 2000.
- [26] J. Zhong and S. Sclaroff, "Segmenting foreground objects from a dynamic textured background via a robust Kalman filter," in *Proceedings of the 9th IEEE International Conference on Computer Vision*, pp. 44–50, October 2003.
- [27] Y. Zhang and Y. Liu, "Missing traffic flow data prediction using least squares support vector machines in urban arterial streets," in *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2009*, pp. 76–83, April 2009.

*Research Article*

# **Modeling the Joint Choice Decisions on Urban Shopping Destination and Travel-to-Shop Mode: A Comparative Study of Different Structures**

**Chuan Ding, Binglei Xie, Yaowu Wang, and Yaoyu Lin**

*Shenzhen Key Laboratory of Urban Planning and Decision Making Simulation, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China*

Correspondence should be addressed to Binglei Xie; xiebinglei@163.com

Received 5 November 2013; Revised 10 February 2014; Accepted 12 February 2014; Published 13 March 2014

Academic Editor: Huimin Niu

Copyright © 2014 Chuan Ding et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The joint choice of shopping destination and travel-to-shop mode in downtown area is described by making use of the cross-nested logit (CNL) model structure that allows for potential interalternative correlation along the both choice dimensions. Meanwhile, the traditional multinomial logit (MNL) model and nested logit (NL) model are also formulated, respectively. This study uses the data collected in the downtown areas of Maryland-Washington, D.C. region, for shopping trips, considering household, individual, land use, and travel related characteristics. The results of the model reveal the significant influencing factors on joint choice travel behavior between shopping destination and travel mode. A comparison of the different models shows that the proposed CNL model structure offers significant improvements in capturing unobserved correlations between alternatives over MNL model and NL model. Moreover, a Monte Carlo simulation for a group of scenarios assuming that there is an increase in parking fees in downtown area is undertaken to examine the impact of a change in car travel cost on the joint choice of shopping destination and travel mode switching. The results are expected to give a better understanding on the shopping travel behavior.

## **1. Introduction**

Both of the destination choice and travel mode choice for shopping trips play important roles in travel demand analysis and transportation policy assessment. Consequently, understanding factors influencing travelers' destination and travel mode choice is necessary to examine the potential effectiveness of policy measures. Previous studies have widely focused on the destination choice [1–3] and travel mode choice [4, 5], respectively. In the past few years, research on joint choice travel behavior was focused primarily on the field of travel mode and departure time [6, 7], as well as residential location and travel mode [8, 9]. As many researchers mentioned, there is a strong relationship between shopping destination and travel mode choice, and people also often make the two decisions simultaneously [10]. Joint analysis of shopping destination and travel mode is helpful to understand the interactions between them and is necessary to assess the impact of the transport policies.

Multinomial logit (MNL) model and nested logit (NL) model based on random utility maximization have been most widely used to analyze travel behavior [11]. However, the MNL model imposes the restriction that the distribution of random error terms is independent and identical over alternatives, which leads to the independence of irrelevant alternatives (IIA) property. Therefore, unobserved similarities existing among choice alternatives in MNL model are overlooked. The most widely known relaxation of MNL model is the NL model. For the NL model, a uniform amount of correlation within a nest of alternatives is allowed, but alternatives not located in the same nest are uncorrelated [12]. For the joint choice analysis of shopping destination and travel mode, two appropriate structures based on the NL model can be described: one is used to nest by shopping destination; another is used to nest by travel mode. However, the both structures can only accommodate correlation along one of the two dimensions. In recent years, the CNL model

has received more attentions in the literature, which allows alternatives to belong to more than one nest instead of each alternative being restricted to a single nest in NL model [13, 14]. Therefore, the CNL model has a more flexible correlation structure to account for various patterns of similarity and dissimilarity among alternatives [8, 15].

In summary, shopping destination choice has received relatively less attention than other travel behaviors such as travel mode and departure time [16, 17]. Studies on simultaneous choice analysis of shopping destination and travel mode that allows for the flexible correlations along the both choice dimensions are limited. In addition, most previous studies just only focused on analyzing the influencing factors for the travel behavior of shopping destination and travel mode choice, while simulation approach related to transport policies based on the estimated model is limited.

In this study, the simultaneous choice of shopping destination and travel mode is described by using a new CNL structure that allows for the joint representation of inter-alternative correlation along the both choice dimensions. Traditional MNL model and NL model are also formulated, respectively, and a comprehensive study to compare the different model structures is carried out. Moreover, based on the estimated model, a Monte Carlo simulation for a series of scenarios assuming that there is an increase in parking fees in downtown area is undertaken to examine the effects of a change in car travel cost on the joint choice of shopping destination and travel mode switching.

The remainder of this paper is organized as follows. The next section presents the model structures used in this study. The third section describes the data used for the model and the fourth section presents the model results. In the fifth section a change in car travel cost due to higher parking fees is simulated based on Monte Carlo method. The final section provides a summary and conclusions.

## 2. Model Specification

In contrast to previous studies on shopping destination choice, the shopping destination choice set in this paper is a selection of spatial areas according to the shopping distance from home location rather than a selection of zones. Therefore, in this study, home location is assumed to be exogenous. Shopping destination and travel behavior are concentrated within the downtown area, which is a classic example of the monocentric city. So the trips are generated by residents who dwell in the central business district (CBD). Therefore, the shopping destination choice set is based on a series of concentric road-distance rings around the residence. The shopping distance is measured between residential and shopping destination. Travel time and travel cost are computed as a function of shopping distance, which can be obtained from the Maryland Statewide Transportation Model (MSTM).

The shopping destination subset has 3 alternatives consisting of concentric road-distance rings around residence within 1 mile, 1-2 miles, and over 2 miles. The travel mode

TABLE 1: Alternatives for joint choice of shopping destination and travel-to-shop mode.

Alternatives	Distance to shop from home location	Mode of travel to shop
Alternative 1	Within 1 mile	Car
Alternative 2	Within 1 mile	Transit
Alternative 3	Within 1 mile	Walk and bicycle
Alternative 4	1-2 miles	Car
Alternative 5	1-2 miles	Transit
Alternative 6	1-2 miles	Walk and bicycle
Alternative 7	Over 2 miles	Car
Alternative 8	Over 2 miles	Transit
Alternative 9	Over 2 miles	Walk and bicycle

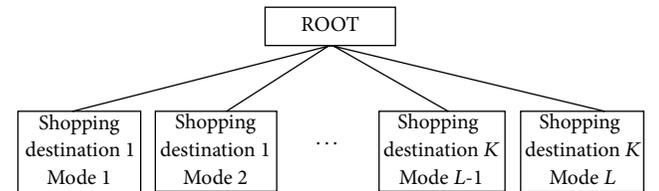


FIGURE 1: Structure of MNL model.

choice subset consists of 3 modes of home-based travel-to-shop: car, transit, walk, and bicycle. Therefore, the model choice set is defined as the joint choice set of shopping destination and travel mode, which creates a set of  $N = 9$  alternatives for each decision-maker located at CBD, as shown in Table 1.

*2.1. Multinomial Logit and Nested Logit Models.* In the past few years, many discrete choice models were developed based on the generalized extreme value (GEV) theory proposed by McFadden [18]. The GEV models are able to capture the unobserved similarities among alternatives, thus relaxing the restriction of MNL and NL models. Several specific GEV models have been formulated by Wen and Koppelman [19] and by Daly and Bierlaire [20]. In this study, all the model structures are presented based on the GEV model framework to analyze the joint choice behavior of shopping destination and travel mode, in order to capture the unobserved correlations between alternatives.

The basic structure tested is an MNL model assuming that no correlations exist between any of the alternatives. The nesting structure is shown in Figure 1.

There are two possible two-level NL structures based on nesting different dimensions of the choice. For example, one appropriate NL structure for the two-level combined model based on nesting by shopping destination is shown in Figure 2, with shopping destination at the upper level and travel mode at the lower level. Alternatives are grouped together based on the shopping destination dimension. In the two-level NL structure, each nest has its own nesting parameter  $\mu$  ( $0 < \mu \leq 1$ ). The nesting parameter can be used to capture the correlations between alternatives sharing

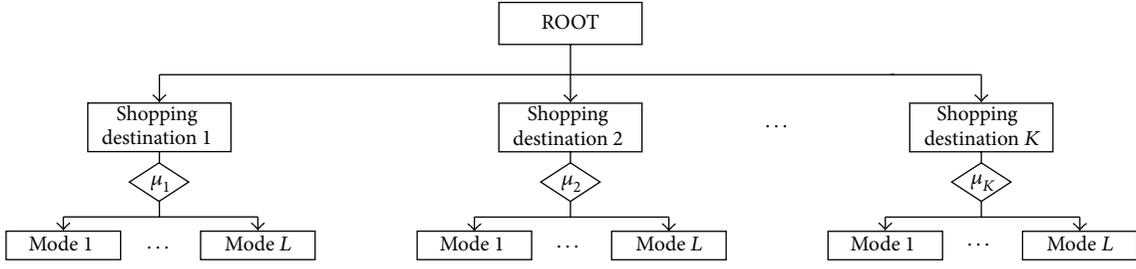


FIGURE 2: Structure of two-level NL model, using nesting along shopping destination dimension.

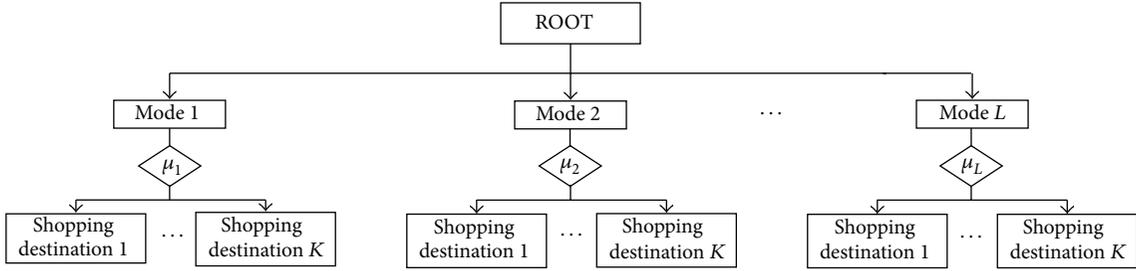


FIGURE 3: Structure of two-level NL model, using nesting along travel mode dimension.

the nest of shopping destination. It is also called dissimilarity parameter. The correlation between alternatives sharing the same nest increases as the dissimilarity parameter decreases.

Figure 3 shows another nesting structure that alternatives are grouped together based on the travel mode dimension. The NL structure shown in Figures 2 and 3 can only be used to analyze the correlation along only one of the two dimensions. They cannot be used to analyze the correlations along the two dimensions of choice simultaneously. For example, the model structure shown in Figure 2 cannot be used to capture spatial correlation between alternative using mode  $L$  to destination  $K_1$  and the alternative using mode  $L$  to destination  $K_2$ . In general, if there are  $K$  dimensions in the choice process, joint choice NL model used in most previous studies can only be used to analyze the correlations along at most  $K - 1$  of  $K$  dimensions by using a multilevel structure.

**2.2. Cross-Nested Logit Model.** The deficiencies of the MNL and NL model structures were first discussed by Hess and Polak in the context for air travel behavior [15]. The solution put forward by Hess and Polak is to use a CNL model structure. It is one motivation for the efforts made in this study to propose improved structures for the joint choice of shopping destination and travel mode. Based on the previous studies, a new CNL model structure is proposed in Figure 4. As shown in Figure 4, the structure for the joint choice model is specified by allowing each alternative to belong to exactly one nest in each shopping destination and travel mode groups. As such, the structure of the model is able to accommodate full correlations along all the dimensions using the simultaneous pattern. In this paper, the allocation parameters  $\alpha$  ( $0 \leq \alpha \leq 1$ ), governing the proportion by which an alternative belongs to each nest, can also be obtained based on the GEV structure. A value of zero

indicates that the alternative does not belong to the nest at all. It is usually specified that the allocation parameters for a given alternative must sum to unity over all nests. In this study, the nonzero allocation parameters for a given alternative were fixed to a value of 0.5, indicating that an alternative belongs by the same proportion to one shopping destination nest and one travel mode nest. As such, the improved structure of the model is able to accommodate the correlations between alternatives along all the dimensions using the simultaneous pattern.

**2.3. Model Formulation.** As a specific GEV model, the CNL model is formulated for the joint probability choice of shopping destination and home-based travel-to-shop mode. There are two main advantages for the application of the CNL structure. On one hand, the CNL model structure provides a more flexible correlation structure of the error term that allows the potential correlations between alternatives to be captured along the both choice dimensions. On the other hand, the CNL mode structure has closed-form expression derived for the calculation of the choice probability.

According to the GEV theorem [18, 19, 21], the CNL model choice probability derived from the generator function presented in (1) is defined in terms of conditional and marginal probabilities as shown in (2):

$$G(y) = \sum_m \left( \sum_k (\alpha_m y_k)^{1/\mu_m} \right)^{\mu_m}, \quad (1)$$

$$P(k) = \sum_m P(k | m) P(m), \quad (2)$$

where  $k$  represents an alternative;  $m$  represents a nest;  $\mu_m$  is a nest-specific coefficient;  $\alpha_m$  is a weight parameter;

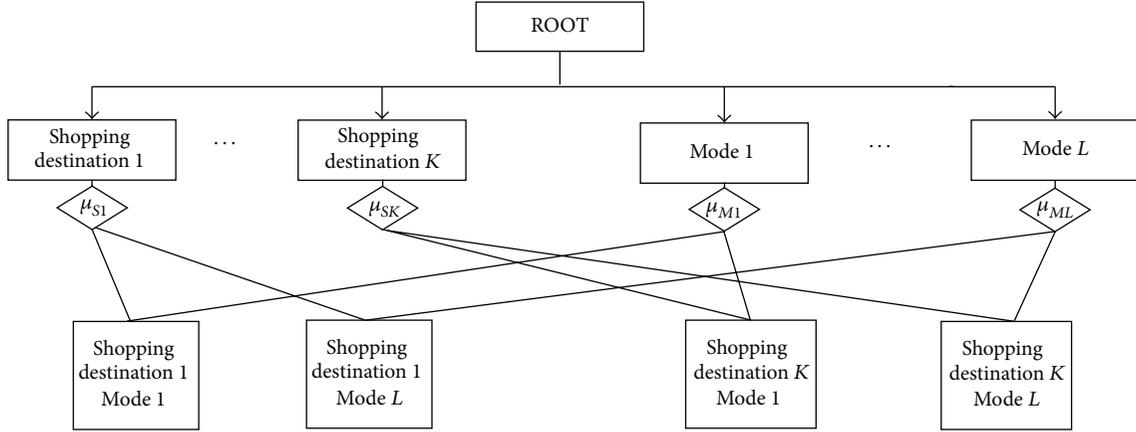


FIGURE 4: Structure of two-level CNL model, using nesting along the shopping destination and travel mode.

the conditional probability of an alternative  $k$  being chosen in nest  $m$  is as follows:

$$P(k | m) = \frac{(\alpha_{mk} e^{V_k})^{1/\mu_m}}{\sum_k (\alpha_{mk} e^{V_k})^{1/\mu_m}}, \quad (3)$$

where  $\alpha_{mk}$  is an allocation parameter that characterized the portion of alternative  $k$  assigned to nest  $m$ ,  $0 \leq \alpha_{mk} \leq 1$ . And the marginal probability of a nest  $m$  being chosen is shown as follows:

$$P(m) = \frac{\left(\sum_k (\alpha_{mk} e^{V_k})^{1/\mu_m}\right)^{\mu_m}}{\sum_m \left(\sum_k (\alpha_{mk} e^{V_k})^{1/\mu_m}\right)^{\mu_m}}. \quad (4)$$

Thus, the probability of the CNL alternative  $k$  being chosen is shown as follows:

$$\begin{aligned} P(k) &= \sum_m P(k | m) P(m) \\ &= \sum_m \left( \frac{(\alpha_{mk} e^{V_k})^{1/\mu_m}}{\sum_k (\alpha_{mk} e^{V_k})^{1/\mu_m}} \cdot \frac{\left(\sum_k (\alpha_{mk} e^{V_k})^{1/\mu_m}\right)^{\mu_m}}{\sum_m \left(\sum_k (\alpha_{mk} e^{V_k})^{1/\mu_m}\right)^{\mu_m}} \right). \end{aligned} \quad (5)$$

In (5), there are two key factors on which the probability of the alternative  $k$  choosing depends: nesting coefficients  $\mu_m$  and deterministic component  $V_k$  of the utility function. In this study, the parameters are estimated based on maximum likelihood method.

### 3. Data Sources and Sample Formation

The data used in this study is drawn from the Baltimore and Washington regional household travel survey (HTS), which was conducted by Baltimore Metropolitan Council (BMC) and Transportation Planning Board at the Metropolitan Washington Council of Governments (MWCOC) during 2007-2008. The areas selected for this analysis is the Baltimore

City and Washington, D.C., which are the downtown areas of the Maryland-Washington, D.C. region. In addition to the HTS dataset, there are other three important sources of data used in the analysis: origin-destination travel time and cost matrices by different modes from Maryland Statewide Transportation Model (MSTM); digital data for GIS analysis provided by National Center for Smart Growth (NCSG) at the University of Maryland; and land use and employment data in traffic analysis zones (TAZs) from Metropolitan Planning Organizations (MPOs) and Quarterly Census Employment and Wages (QCEW).

Many factors have been identified that influence the decisions of shopping destination and travel mode [8, 22, 23]. There are four variable groups used in this analysis: household, individual, land use related, and travel related characteristics. The variables of household characteristics include household size, income, and the number of cars available in the household. The variables of individual characteristics include gender and age. The built environment at the home-located TAZ is found to be potentially important variables influencing the choice of shopping destination in many previous studies. In this study, population density and retail employment density at the TAZ level are used as land use related explanatory variables. Travel related characteristics include travel time and travel cost computed from home location to shopping destination by different travel modes. The total variables used in analysis are shown in Table 2.

The distributions of shopping distance and travel time for all trips are shown in Figures 5 and 6. The distribution of shopping distance shows that the shopping trips decrease as the distance from home increases, which is consistent with expectations. Most shoppers tend to make shopping trips within one mile distance from home. Figure 6 shows that most shoppers tend to take less than twenty minutes for their shopping trips.

A descriptive analysis is conducted to get intuitive findings regarding the association between household, individual, land use related characteristics and the preferences of shopping destination and travel mode. As shown in Table 3,

TABLE 2: Descriptive statistics of the sample data for home-based travel-to-shop trips ( $N = 975$ ).

Variable name	Variable description	Mean	St. Dev.
Household characteristics			
Household size	Single person household (1 = yes; 0 = otherwise)	0.34	0.475
	Household size is equal to two persons (1 = yes; 0 = otherwise)	0.44	0.496
	Household size is equal to or more than three persons (1 = yes; 0 = otherwise)	0.22	0.414
Household income	Household income is less than \$30,000 (1 = yes; 0 = otherwise)	0.23	0.422
	Household income is between \$30,000 and \$100,000 (1 = yes; 0 = otherwise)	0.42	0.494
	Household income is equal to or more than \$100,000 (1 = yes; 0 = otherwise)	0.35	0.477
Cars ownership	Household owns no car (1 = yes; 0 = otherwise)	0.22	0.415
	Household owns one car (1 = yes; 0 = otherwise)	0.42	0.494
	Household owns two or more cars (1 = yes; 0 = otherwise)	0.36	0.479
Individual characteristics			
Gender	Male (1 = yes; 0 = otherwise)	0.45	0.498
Age	Person is less than 25 years old (1 = yes; 0 = otherwise)	0.05	0.210
	Person is between 25 and 54 years old (1 = yes; 0 = otherwise)	0.52	0.500
	Person is equal to or more than 55 years old (1 = yes; 0 = otherwise)	0.44	0.496
Land use related characteristics			
Residential density	Continuous variable: residential density of TAZ in which home lies (people/acre)	25.37	18.049
Retail employment density	Continuous variable: retail employment density of TAZ in which home lies (retail employment/acre)	2.98	5.878
Travel related characteristics			
Travel Time	Continuous variable: total time of a trip (min) provided by MSTM		
Travel cost	Continuous variable: total travel cost for a trip (\$) as a function of distance provided by MSTM		
Choice characteristics			
Shopping destination	Distance to shop: within 1 mile, 1-2 miles, and over 2 miles		
Travel mode	Trip modes: car, transit, walk, and bicycle		

TABLE 3: Sample profiles stratified by shopping destination and travel mode ( $N = 975$ ).

Frequencies (%)	Shopping destination			Travel mode		
	Within 1 mile	1-2 miles	Over 2 miles	Car	Transit	Walk and bicycle
Household size						
Single person	44.6	14.6	40.8	48.2	27.6	34.2
Two persons	36.2	17.9	45.9	59.1	12.2	28.7
More than two persons	42.1	19.2	38.8	69.6	6.1	24.3
Household income						
<\$30,000	46.9	14.2	38.9	41.6	25.7	32.7
\$30,000–\$100,000	38.1	17.6	44.3	62.8	9.8	27.4
>\$100,000	38.8	18.2	42.9	62.1	7.6	30.3
Cars ownership						
No car	56.3	12.1	31.6	7.4	38.1	54.4
One car	40.2	18.6	41.2	64.2	6.8	29.1
More than one car	30.8	18.2	51.0	81.0	4.0	15.0
Gender						
Male	42.4	18.1	39.5	55.6	12.7	31.7
Female	38.8	16.1	45.1	59.4	12.7	27.9
Age						
16–24	33.3	11.1	55.6	57.8	8.9	33.3
25–54	42.7	16.1	41.2	52.5	12.9	34.6
>55	38.4	18.7	42.9	63.7	12.9	23.4

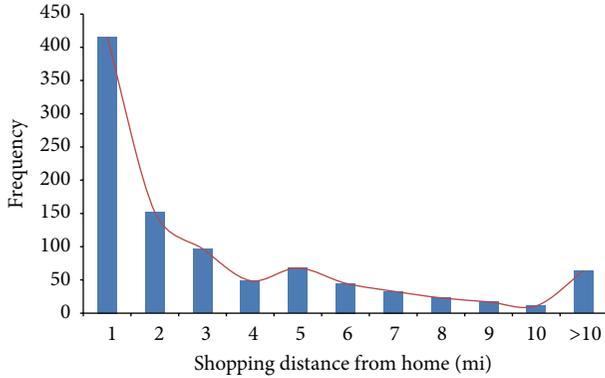


FIGURE 5: Distribution of shopping distance for the home-based travel-to-shop trips.

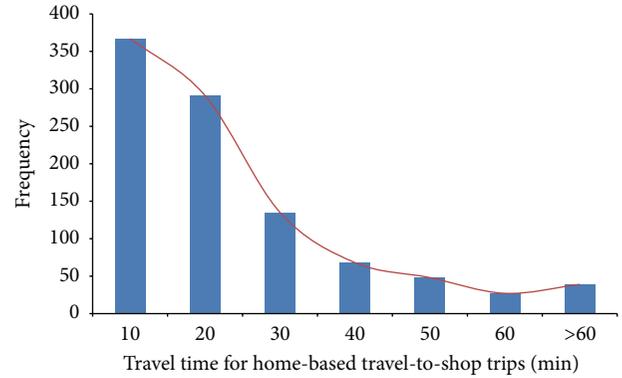


FIGURE 6: Distribution of travel time for the home-based travel-to-shop trips.

young individuals are more likely to shop farther away from home and they are also found to show a negative propensity to use transit. This may be seen as an intuitive character-related effect for young individuals. As expected, people from smaller household size, lower household income, and lower car ownership are found to be more likely to shop closer from home and use transit or walk and bicycle for shopping trips.

#### 4. Empirical Results

All the models presented in this study were estimated using Biogeme [24, 25], including the MNL model, two kinds of NL model, and the proposed CNL model. The probability of choosing each alternative can be estimated using the presented model based on the given independent variables.

The travel-related parameters, data fit measures, and dissimilarity parameters for the four models are presented in Table 4. In terms of adjusted  $\rho^2$ , it can be seen that the MNL model and the first NL model outperform others. In terms of the log-likelihood, the final log-likelihood value of the CNL is  $-1307.762$ , which is  $0.846$ ,  $0.509$ , and  $56.683$  points higher, respectively, than that of the MNL model and the other two kinds of NL model. As expected, the signs of the travel-related parameters are negative. The average value of travel time savings for the shopping trips is about  $0.19$  \$/min (about  $11.5$  \$/hour), which is lower than the value for commuting trips reported by Hess et al. [26].

In terms of the unobservable correlations between alternatives, as seen from Table 4, the CNL model is superior in capturing the unobservable correlation between alternatives when compared to the MNL and the other two kinds of NL model. The dissimilarity parameter along the transit dimension is minimal, indicating that the alternatives in the transit nest have high correlations. In other words, the dissimilarity parameters capture the pattern of substitutability across alternatives [19, 27]. Due to the high substitutability of the alternatives in the transit nest, the decision-makers are more likely to shift their shopping destination rather than their travel mode when the values of the utility variables change (such as due to transportation control measures).

The mean value of the dissimilarity parameter for the shopping travel mode is lower than that for shopping destination, which means that the shoppers who live in the downtown area are more likely to shift their travel mode than shopping destination.

The detailed estimation results based on the CNL model are presented in Table 5. The model results suggest that these household, individual, and land use characteristics in the case study area are the important factors influencing the individuals' shopping destination and travel mode choice decisions. In terms of the household characteristics, it can be seen from Table 5 that single person is significantly less likely to shop by transit when the shopping destination is 1-2 miles away home, compared with the base alternative. People from the larger households are significantly more likely to shop closer from home and significantly less likely to walk to shop. Low income groups are found to choose shopping destinations closer from home using transit, compared with the base alternative. However, they are found significantly less likely to walk to shop. Higher income groups show a positive propensity to walk to shop, compared with the base alternative, though it is less significant at the 95 percent level. As expected, low car ownership level is found to show a significantly negative propensity to drive to shop. People with more cars available are significantly less likely to walk to shop even if the shopping destination is within 1 mile from home. In terms of the individual characteristics, the variable of gender is found to be less significant. Young adults are significant more likely to walk to the shopping destination further away from home. Older individuals are found to be more likely to drive to shop than younger individuals when the shopping destination is within 1 mile from home. In terms of the land use related characteristics, it is found that people who live in high residential density and retail employment density areas are significantly more likely to walk to shop within 1 mile from home.

#### 5. Monte Carlo Simulation

The traffic problem becomes much serious in the downtown areas. Therefore, the simulation tests on different scenarios

TABLE 4: Travel related parameter and data fit measures ( $N = 975$ ).

Attributes, dissimilarity parameter, and data fit measures	MNL		NL using nesting by shopping destination		NL using nesting by travel mode		CNL using nesting by shopping destination and travel mode	
	Parameter	$t$ -stat	Parameter	$t$ -stat	Parameter	$t$ -stat	Parameter	$t$ -stat
Travel cost	-0.392	-6.42	-0.392	-6.39	-0.346	-6.60	-0.389	-6.36
Travel time	-0.0754	-9.52	-0.0755	-9.51	-0.0643	-9.81	-0.0754	-9.66
$\mu_{\text{destination1}}$	1.00	—	1.00	—	—	—	1.06	0.10
$\mu_{\text{destination2}}$	1.00	—	1.00	6.49	—	—	1.62	0.30
$\mu_{\text{destination3}}$	1.00	—	1.00	4.96	—	—	1.48	1.16
$\mu_{\text{car}}$	1.00	—	—	—	1.00	—	1.00	0.03
$\mu_{\text{transit}}$	1.00	—	—	—	2.04	7.57	1.02	0.16
$\mu_{\text{walk}}$	1.00	—	—	—	1.26	3.65	1.00	0.04
VTTs (\$/min)	0.1923		0.1926		0.1858		0.1938	
Final LL	-1308.608		-1308.271		-1364.445		-1307.762	
Adjusted $\rho^2$	0.343		0.343		0.316		0.333	
Parameters	98		100		100		122	

Note. Alternative 1 is the reference category; LL: log-likelihood; VTTs: value of travel time savings.

are extremely useful for the transportation demand management (TDM), transportation control measures (TCM), and intelligent transportation system (ITS). In this study, another important motivation lies in obtaining the simulated results when the travel-related attributes change arising from transport policies, using the empirical results to test the impact of a change in travel cost on the joint choice of the shopping destination and travel mode switching.

Most transportation congestion management actions attempt to affect the mode choice behavior or reduce trip making by directly or indirectly impacting the level-of-service variables. For example, congestion pricing and parking fees rely on the use of monetary disincentives for the car mode. In this study, a group of simulations is carried out by assuming that there is an increase in car travel cost due to the higher parking fees in the downtown area.

Sample enumeration is used to calculate the joint choice probabilities for each shopper based on the estimated parameters presented in Table 5. This is extremely useful for producing the aggregate shares for all alternatives. To produce the analysis of the impact of a change in travel-related attributes, the simulated choices following the change can be obtained based on the Monte Carlo simulation using the estimated model. Then the correct predicted probabilities for all alternatives can be calculated based on the simulated each choice. It is found that the predicted shares are very close to the actual shares, as shown in Table 6. Therefore, the CNL model can be used to accurately represent the choice shares in the study area.

The simulated results for one dollar, two and one half dollars, and five dollars increasing in car travel cost due to higher parking fees are presented in Table 7. As expected, the simulated results show that the choice probability of driving decreases with the car travel cost increasing. Specially, the effect of higher parking fees is more significant for the long shopping distance. The shares of using car to far away

from home (i.e., over 2 miles) for shopping sharply decrease when there is a higher parking fees. Most shoppers living in the downtown area will shift from car mode to walk or bicycling for shopping trips to reduce their transport spending. In this case, it is important to provide a suitable walking environment and provide a better neighborhood design for the pedestrian. Otherwise, the people will still choose the car mode to shop because of the bad pedestrian environment, and the policy of a change in parking fees will fail.

## 6. Conclusion

In this study, the joint choice of shopping destination and travel-to-shop mode is analyzed, using three different types of GEV structures: MNL model, two types of the NL, and a new CNL model. A combination of data sources is used to estimate the choice of models for the downtown areas in Maryland-Washington, D.C. region. As the estimated results showed, unobserved similarities which exist among choice alternatives are overlooked in the MNL model, the use of two-level NL models can allow for the treatment of correlation along a single dimension of choice, and the proposed CNL model can capture the unobserved correlations along the both shopping destination and travel mode dimensions. In terms of model performance, the CNL model outperforms other models in general. Therefore, the CNL model can be seen as a valuable tool in the analysis of the joint choice of shopping destination and travel mode. The model results show that household, individual, land use, and travel related characteristics play different roles in the joint choice behavior of shopping destination and travel mode.

A series of simulations are conducted for increasing car travel cost to forecast the aggregate choice shares, using the sample enumeration method. Significant choice switching effects are found, and the simulated results suggest that

TABLE 5: Estimation results of the CNL model.

Variables	Within 1 mile			1-2 miles			Over 2 miles				
	Transit Parameter	Walk and bicycle Parameter	Car Parameter	Transit Parameter	Walk and bicycle Parameter	Car Parameter	Transit Parameter	Walk and bicycle Parameter	Car Parameter		
Size1	-0.699	-0.292	-0.84	-0.15	-1.040	-0.191	-0.64	-0.183	-0.44	0.052	0.06
Size3	-0.328	-0.676	-2.04**	-0.60	-1.400	-0.796	-2.92***	-1.480	-2.86***	-1.010	-1.20
Income1	2.060	-1.170	-3.00***	-1.02	-1.540	-0.324	-1.04	-0.633	-1.37	0.360	0.36
Income3	-2.140	0.388	1.27	0.0620	0.906	-0.007	-0.03	0.192	0.46	1.810	1.91*
Car1	4.030	4.08***	6.21***	-0.66	3.680	0.153	0.25	4.390	6.84***	2.930	2.92***
Car3	-2.180	-0.887	-2.82***	-0.84	-0.699	-0.043	-0.17	-0.227	-0.50	-0.956	-1.12
Household characteristics											
Individual characteristics											
Gender	-0.792	-1.17	-0.89	-0.51	0.148	-0.248	-1.14	-0.349	-1.07	-1.220	-1.68*
Age1	-1.060	1.000	1.35	0.88	-1.260	0.884	1.35	1.420	1.61	3.160	3.02***
Age3	0.477	0.72	-1.030	-0.07	-0.930	-0.447	-1.93*	-0.641	-1.87*	-1.160	-1.53
Land use related characteristics											
Population density	-0.017	0.014	1.96**	-1.62	0.025	-0.0118	-1.75*	0.006	0.67	-0.016	-0.66
Retail employment density	0.082	1.30	2.71***	-0.004	0.08	-0.015	-0.50	0.038	1.11	-0.064	-0.54

Note. Alternative 1 is the base alternative. \*\*\* indicates significance at the 99 percent level; \*\* at the 95 percent level; \* at the 90 percent level.

TABLE 6: Comparisons between actual shares and predicted shares using sample enumeration.

Actual/predicted shares (%)	Within 1 mile			1-2 miles			Over 2 miles		
	Car	Transit	Walk and bicycle	Car	Transit	Walk and bicycle	Car	Transit	Walk and bicycle
Actual shares	12.9	1.5	25.9	12.0	2.6	2.5	32.7	8.6	1.2
Predicted shares	12.4	1.7	25.6	12.2	2.6	2.3	33.4	8.5	1.4

TABLE 7: Predicted shares based on different scenarios.

Scenario group (%)	Within 1 mile			1-2 miles			Over 2 miles		
	Car	Transit	Walk and bicycle	Car	Transit	Walk and bicycle	Car	Transit	Walk and bicycle
Base scenario	12.4	1.7	25.6	12.2	2.6	2.3	33.4	8.5	1.4
One dollar	11.5	1.6	28.8	11.0	2.9	2.5	30.6	9.3	1.6
Two and one half dollars	9.7	2.1	33.4	9.4	3.3	2.9	25.8	11.4	1.9
Five dollars	6.8	2.2	41.5	6.4	4.0	3.6	17.8	15.3	2.4

transport policies aimed at reducing traffic congestion in downtown areas by increasing car travel cost may have better effects; however improving the pedestrian environment is also necessary.

Charging for the road user has been seen by some academics and urban planners as a solution to traffic problems in the city. The framework presented in this paper has more potential application in the future, such as the impact study of a hypothetical road user charging scheme and the effect analysis of parking charges, transit subsidies, and flexible work hours on a traveler’s behavior. Further studies will not only include the application of the framework based on the CNL model but also include the use of advanced model structures allowing joint for the cross-nesting, continuous deterministic, and random taste heterogeneity to exam the shopper travel behavior.

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgments**

This research is supported in part by the National Natural Science Foundation of China (no. 71173061), Shenzhen Science and Technology Development Funding-Fundamental Research Plan (Grant JCYJ20120615145601342), and Key Laboratory of Eco Planning & Green Building, Ministry of Education (Tsinghua University, 2013U-6).

**References**

[1] W. W. Recker and L. P. Kostyniuk, “Factors influencing destination choice for the urban grocery shopping trip,” *Transportation*, vol. 7, no. 1, pp. 19–33, 1978.

[2] C. R. Bhat and J. Guo, “A mixed spatially correlated logit model: formulation and application to residential choice modeling,” *Transportation Research Part B*, vol. 38, no. 2, pp. 147–168, 2004.

[3] A. Sivakumar and C. R. Bhat, “Comprehensive, unified framework for analyzing spatial location choice,” *Transportation Research Record*, no. 2003, pp. 103–111, 2007.

[4] C. R. Bhat, “Work travel mode choice and number of non-work commute stops,” *Transportation Research Part B*, vol. 31, no. 1, pp. 41–54, 1997.

[5] A. de Palma and D. Rochat, “Mode choices for trips to work in Geneva: an empirical analysis,” *Journal of Transport Geography*, vol. 8, no. 1, pp. 43–51, 2000.

[6] C. A. Tringides, X. Ye, and R. M. Pendyala, “Departure-time choice and mode choice for nonwork trips alternative formulations of joint model systems,” *Transportation Research Record*, no. 1898, pp. 1–9, 2004.

[7] S. Bajwa, S. Bekhor, M. Kuwahara, and E. Chung, “Discrete choice modeling of combined mode and departure time,” *Transportmetrica*, vol. 4, no. 2, pp. 155–177, 2008.

[8] A. Vega and A. Reynolds-Feighan, “A methodological framework for the study of residential location and travel-to-work mode choice under central and suburban employment destination patterns,” *Transportation Research Part A*, vol. 43, no. 4, pp. 401–419, 2009.

[9] L. Yang, G. Zheng, and X. Zhu, “Cross-nested logit model for the joint choice of residential location, travel mode, and departure time,” *Habitat International*, vol. 38, pp. 157–166, 2013.

[10] J. Scheiner, “Interrelations between travel mode choice and trip distance: trends in Germany 1976–2002,” *Journal of Transport Geography*, vol. 18, no. 1, pp. 75–84, 2010.

[11] G. de Jong, A. Daly, M. Pieters, C. Vellay, M. Bradley, and F. Hofman, “A model for time of day and mode choice using error components logit,” *Transportation Research Part E*, vol. 39, no. 3, pp. 245–268, 2003.

[12] S. Hess, M. Fowler, T. Adler, and A. Bahreinian, “A joint model for vehicle type and fuel type choice: evidence from a cross-nested logit study,” *Transportation*, vol. 39, no. 3, pp. 593–625, 2012.

[13] A. Papola, “Some developments on the cross-nested logit model,” *Transportation Research Part B*, vol. 38, no. 9, pp. 833–851, 2004.

[14] M. Bierlaire, “A theoretical analysis of the cross-nested logit model,” *Annals of Operations Research*, vol. 144, no. 1, pp. 287–300, 2006.

[15] S. Hess and J. W. Polak, “Exploring the potential for cross-nesting structures in airport-choice analysis: a case-study of the Greater London area,” *Transportation Research Part E*, vol. 42, no. 2, pp. 63–81, 2006.

[16] M. D. Abkowitz, “An analysis of the commuter departure time decision,” *Transportation*, vol. 10, no. 3, pp. 283–297, 1981.

- [17] K. A. Small, "The scheduling of consumer activities: work trips," *American Economic Review*, vol. 72, no. 3, pp. 467–479, 1982.
- [18] D. McFadden, "Modeling the choice of residential location," in *Spatial Interaction Theory and Residential Location*, A. Karlqvist, Ed., pp. 75–96, North-Holland, Amsterdam, The Netherlands, 1978.
- [19] C.-H. Wen and F. S. Koppelman, "The generalized nested logit model," *Transportation Research Part B*, vol. 35, no. 7, pp. 627–641, 2001.
- [20] A. Daly and M. Bierlaire, "A general and operational representation of generalised extreme value models," *Transportation Research Part B*, vol. 40, no. 4, pp. 285–305, 2006.
- [21] S. Bekhor and J. N. Prashker, "GEV-based destination choice models that account for unobserved similarities among alternatives," *Transportation Research Part B*, vol. 42, no. 3, pp. 243–262, 2008.
- [22] R. Cervero, "Built environments and mode choice: toward a normative framework," *Transportation Research Part D*, vol. 7, no. 4, pp. 265–284, 2002.
- [23] R. Ewing and R. Cervero, "Travel and the built environment," *Journal of the American Planning Association*, vol. 76, no. 3, pp. 265–294, 2010.
- [24] M. Bierlaire, "The network of GEV model," in *Proceeding of the 2nd Swiss Transport Research Conference*, Verita, Switzerland, 2002.
- [25] M. Bierlaire, "BIOGEME: a free package for the estimation of discrete choice models," in *Proceeding of the 3rd Swiss Transport Research Conference*, Monte Verita, Switzerland, 2003.
- [26] S. Hess, J. M. Rose, and D. A. Hensher, "Asymmetric preference formation in willingness to pay estimates in discrete choice models," *Transportation Research Part E*, vol. 44, no. 5, pp. 847–863, 2008.
- [27] E. Pels, N. Njegovan, and C. Behrens, "Low-cost airlines and airport competition," *Transportation Research Part E*, vol. 45, no. 2, pp. 335–344, 2009.

## Research Article

# Effectiveness of Variable Message Signs on Driving Behavior Based on a Driving Simulation Experiment

**Xuedong Yan and Jiawei Wu**

*MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Xuedong Yan; [xdyan@bjtu.edu.cn](mailto:xdyan@bjtu.edu.cn)

Received 14 November 2013; Accepted 24 January 2014; Published 13 March 2014

Academic Editor: Wuhong Wang

Copyright © 2014 X. Yan and J. Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Variable message signs (VMSs), as one of the important ITS devices, provide real-time traffic information of road network to drivers in order to improve route choice and relieve the traffic congestion. In this study, the effectiveness of VMS on driving behavior was tested based on a driving simulation experiment. A road network with three levels of VMS location to route-diverging intersection and three types of VMS information format was designed in a high fidelity driving simulator platform. Fifty-two subjects who were classified by driver age, gender, and vocation successfully completed this experiment. The experimental results showed that driver characteristics, VMS location, and information format profoundly influence driving behaviors. Based on the research findings, it is suggested that VMS would be positioned between 150 m and 200 m upstream of the diverging point to balance the VMS effects on traffic safety and operation and the graphic information VMS format is better than the format with text message only.

## 1. Introduction

With the rapid development of urbanization and motorization, China has become the world's second largest car country and the constantly climbing number of vehicles in urban road networks leads to more and more serious traffic congestion problems. The traditional countermeasures of traffic congestion alleviation include constructing new roadways, adding new traffic facilities, and strengthening traffic management [1]. While the effects of the traditional methods are limited in the cities with high vehicle ownership, the technologies of intelligent traffic systems (ITS) have been widely applied to solve traffic issues [2].

Variable message signs, as an advanced traffic guidance system, can provide real-time traffic information in urban road networks to help drivers choose the routes with lower traffic volumes. Thus, the vehicles can be distributed reasonably in road networks so as to improve the performance of traffic system [3, 4]. With the assistance of VMSs, the road capacity can be increased noticeably and the utilization efficiency of downstream off-ramp can be enhanced [5]. However, VMS effectiveness is dependent on drivers' route choice behavior, and VMS design and position may influence

lane changing and speed control behaviors. The behaviors represent the drivers' perception to the guidance information and the trust degree to the information [6]. Therefore, it is critical to investigate how the interface type, sign position, and guidance information formats of VMS affect driving behaviors in urban road networks.

A number of previous studies focusing on VMSs and relevant driving behaviors have been conducted. Two typical research methods, questionnaire, and computer simulation experiment have been applied for analyzing main factors that influence drivers' route choices in the VMS environments. Through the questionnaire, it was found that whether drivers accept an item of VMS advice or not is closely associated with drivers' characteristics and their familiarity degree to the road network [7, 8]. Generally, the acceptance rates of drivers who are unfamiliar with the road network are higher than the familiar drivers because the familiar drivers may select driving route based on experience rather than guidance information only [8, 9]. The drivers' personal attributes, such as age, gender, and driving age can also significantly affect the driving behaviors in discovering, understanding, and complying with the guidance information as the travel characteristics of selected routes change [10–13]. Similar to the survey

TABLE 1: Gender and vocation distribution of recruited subjects.

Gender	Professional	Unprofessional	Total
Male	16	14	30 (57.7%)
Female	9	13	22 (42.3%)
Total	25 (48.1%)	27 (51.9%)	52 (100%)

results from questionnaire, the study based on the computer simulation experiment also showed that the factors including characteristics of guidance information and attributes of drivers would affect drivers' route choice [14]. Additionally, the techniques of questionnaire and simulation experiment were also used to optimize VMS design in the road network. It was found that VMSs' information presentation formats, including numerical format, the text description format, and switch-on-light format, may lead to drivers having different comprehension on traffic status of road networks [15]. The computer simulation experiment result indicated that if VMS releases a clear suggestion strategy of route choice, the rate of drivers' choosing suggested routes is higher than that without strategy [16]. Furthermore, Srinivasan and Jocanis conducted computer simulation experiments to investigate drivers' attention demands of four types of turn-by-turn route guidance displays. The results showed that the most effective and preferred VMS display was the one using countdown bars representing distance from the turn location to supplement text message [17].

While the statement preference (SP) methods of questionnaire and computer simulation experiment are limited to provide detailed driving behavior data, driving simulators were applied to investigate drivers' speed control, lane change, and response to VMSs in virtual reality road environments. Driving simulators can provide a well-controlled experimental condition to compare the drivers' behaviors in road networks with different VMS settings. Another advantage of using driving simulator is that it can collect the data which are difficult to achieve in the real world, especially the vehicle's instantaneous velocity [18]. Based on driving simulator experiments, Kolisetty et al. used a driving simulator to study driving speed behavior under VMS and found that VMSs had an effect on subjects with average speed changes from  $-2$  to  $15$  km/h with VMSs compared to without [19]; Jamson et al. evaluated the effects of VMS on driver performance and safety, who pointed out that drivers needed to reduce their speed in order to read four-line monolingual and four-line bilingual text-based signs [20]. Lee and Abdel-Aty designed 24 scenarios in a driving simulator and showed that variable speed limits combined with VMSs were beneficial in reducing speed variation and lowering congestion [21]. Meanwhile, they also found age difference in the reaction to VMSs in the experiments. Erke et al. observed that lane changing behavior happened more frequently in the road network with VMSs than the network without VMSs [22].

Although a few of previous studies involved drivers' behaviors in the VMS traffic environment, there is a lack of research focusing on investigating how VMS position and information format affect driving behavior, especially using a high fidelity driving simulator. The main objective of this

paper is to investigate whether and how VMS position and VMS information format impact drivers' behaviors, such as route choice, speed control, and lane changing, based on a high fidelity driving simulation experiment.

## 2. Methodology

*2.1. Participants.* A total of 57 test subjects were recruited in this experiment. Every subject would encounter the VMSs six times and there were 342 samples that reflect the drivers' performances under VMS. The similar scale of sample sizes was also applied in the driving simulator experiments by previous simulator experiment designs [23, 24]. All of the subjects were local drivers in Beijing, China, with at least one year of driving experience and at least 2500 kilometers annual driving distance. Before the experiment, the drivers were tested and proved that they understand the meaning and function of the VMSs applied in Beijing. Finally, 52 drivers successfully finished but five subjects could not complete the experiment due to driving simulation sickness whose data were excluded from collection. According to subjects' vocations, the 52 drivers were classified into two groups, 25 professional drivers and 27 unprofessional drivers. The subject gender distribution was balanced in each vocation group, as shown in Table 1. Each subject was required to test three driving scenarios (it took about an hour to finish the experiment) and was compensated with 500 Chinese RMB for their participation.

*2.2. Apparatus/Equipment.* A high fidelity driving simulator located in MOE Key Laboratory for Urban Transportation Complex Systems of the Beijing Jiaotong University was used in the study, as shown in Figure 1. The driving simulator has a linear motion base capable of operation with one degree of freedom. It is composed of a visual system with 300 degrees of front view and three rear view mirrors, a full-size cabin of Ford Focus with real operation interface, a digital sound simulation system, a vibration system, and the center console. The visual system allows resolution equal to  $1400 \times 1050$  pixels for each channel. The software, including Simvista and Simcreator, is provided for modeling road networks and driving scenarios. The data sampling frequency is up to 60 Hz.

In addition, five cameras are installed inside and outside the cabin to supervise the experimental process. An emergency stop button is installed both inside cabin beside the driver seat and in the front of control desk in order that either subject or researcher can discontinue the experiment immediately in case the subject suffers driving simulation sickness.

*2.3. Road Network and Driving Scenario Design.* To investigate the effectiveness of VMS on driving behavior, a road network was designed, as shown in Figure 2, which was composed of four-lane road segments with 80 kilometers speed limit and signalized intersections. The network can be divided into two parts. The first part was designed for testing how VMS-I and its location (distance to intersection)



FIGURE 1: The driving simulator.

TABLE 2

Scenario	The first part VMS-I location	The second part VMS-II information format
A	0 m	Text-only format
B	200 m	Graphics-only format
C	400 m	Combination of text and graphics

affect drivers' route choice, speed control, and lane changing behaviors. The distance between VMS-I and signalized intersection were set as 0 m, 200 m, and 400 m for three driving scenarios, respectively. In each scenario, VMS-I released the same content in the graphic format (see Figure 2). The graphic format displayed the current congestion state of selectable routes in the network, in which green color represents smooth and red color represents jam. The graphic format of VMS is most frequently applied in Beijing. The second part of the simulated network was designed to investigate the impact of information delivery format of VMS-II on the driving behaviors. There were three kinds of information formats tested in this experiment, including text-only, graphics-only, and combination of text and graphics. In the VMS-II, the text-only format released the information "800 METERS AHEAD - TRAFFIC CONGESTION"; the graphics format displayed the current congestion state of selectable routes, which is the same design as VMS-I; the combination format of text and graphics displayed both the current congestion state of selectable routes and suggestion of route information with text "TRAFFIC CONGESTION - PLEASE DETOUR SUZHOU ROAD." The three formats of VMS-II were located at 200 meters ahead from the intersection and tested in the three driving scenarios, respectively. The three scenarios were coded as A, B, and C, as summarized in Table 2.

In each driving scenario, subjects needed to drive from the start point to the end points, as illustrated in Figure 2. The distance from start point to the VMS-I location is 1500 m, which is sufficient for drivers to accelerate to their normal

driving speeds. Since subjects departed the driving simulator in the inside lane at the start point, if drivers wanted to choose the routes which were not congested, they needed to change from inside lane into outside lane before approaching the intersections. In order to avoid that the signal lights changes disturb subjects' route choice decisions, the signal lights always kept green along each driving route. In addition, to make the driving scenarios seem more realistic, a number of vehicles were design downstream of the intersection, where the segment was displayed as congestion state in the VMS. When subjects approached the intersection, the proximity sensor would trigger the vehicles driving at 5 km/h along the segment, which caused subjects to percept traffic congestion. If subjects selected the congested route, the vehicles would restore to normal driving speed around 80 km/h after traveling one kilometer forward. The design of traffic congestion and driving scenario was shown in Figure 3.

*2.4. Experimental Procedure.* Upon arrival, the subjects were asked to fill out and sign an informed consent form (per IRB). The subjects were then advised to drive and behave as they normally would and to adhere to traffic laws as in real life situations. The subjects were also notified that they could quit the experiment at any time in case of driving simulation sickness or any kind of discomfort. Prior to the formal experiment, drivers were trained for at least 10 min to familiarize with the driving simulator operation and the experimental road network. During the course of the practice, subjects exercised selected maneuvers including straight driving, acceleration, deceleration, left/right turn, and other basic driving behaviors. Then, the formal experiments began during which all subjects would test the three scenarios A, B, and C in a random order so as to eliminate the experimental time order effect. For security and liability reasons, each subject was escorted to the simulator cabin to commence the experiment and he/she was allowed at least 20 min to rest before running the next scenario.

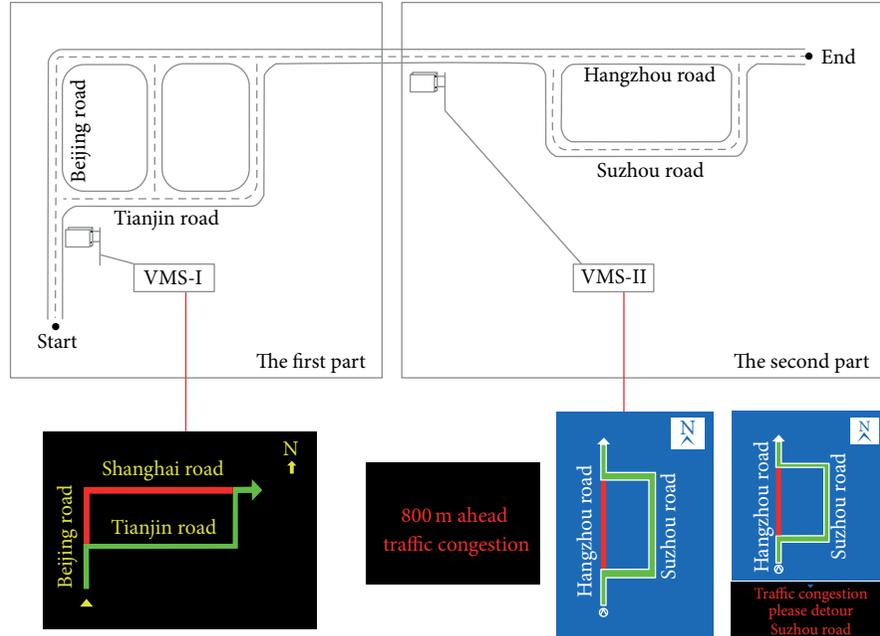


FIGURE 2: The experimental road network.

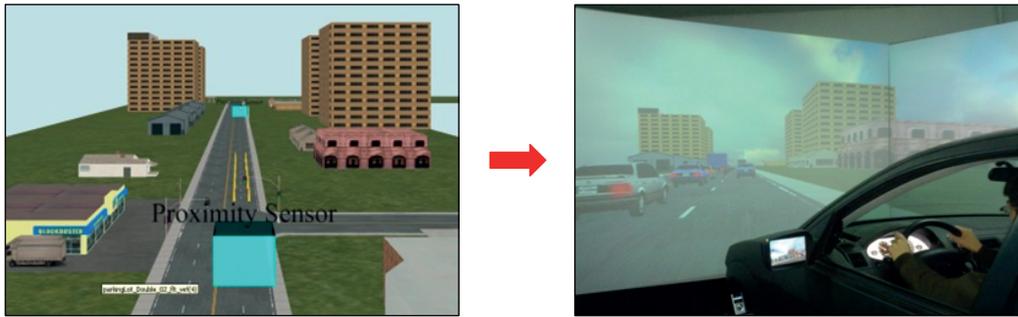


FIGURE 3: The design of traffic congestion and driving scenario.

2.5. *Dependent Measures.* Data collection and analyses were based on each subject driving three times in the simulated road network. Each subject would meet the different types of VMSs six times for a total of 312 route choices. The related dependent measures for driving behavior analyses were defined as follows:

- (i) RC (straight = 0; turn = 1): route choice, whether a driver went straight or turned at an intersection;
- (ii) SPEED (km/h): the vehicle's average speed for every twenty meters upstream or downstream of the VMS;
- (iii) SV (km/h): speed under VMS, the vehicle's operation speed under VMS;
- (iv) LCT (s): lane changing time, which was measured as lane changing duration if a subject has a lane changing behavior;
- (v) LCP (m): lane changing position, which was measured as the distance to the intersection at which a subject started changing lane if a subject has a lane changing behavior;
- (vi) LCL (m): lane changing length, which was measured as the longitudinal distance of lane changing;
- (vii) LCS (km/h): lane changing speed, which was measured as the average speed during lane changing;
- (viii) LCD ( $m/s^2$ ): lane changing deceleration, which was measured as the average deceleration during lane changing.

Based on the driving behavior data, the following results focused on studying the effects of VMS position, VMS information delivery formats, subject vocation, gender, and age on driving behaviors. The hypothesis testing in the following analyses are based on a 0.05 significance level.

### 3. Experiment Results and Discussion

**3.1. Route Choice Analyses.** The binary logistic model is suitable for analyzing route choice behavior because the behavior can be described as a dichotomy variable. The binary logistic regression technique has been applied to explore the relationship between route choice and its potential influencing factors [25–27]. In this study, the parameter estimates of logistic regression models for VMS locations and information formats are shown in Table 3, respectively, where the odds ratios  $\text{Exp}(B)$  in the table is defined as the ratio of turn decision likelihood to keeping-straight likelihood.

In the first part for VMS-I analysis, the regression results indicate that the independent variables of age ( $P < 0.001$ ), gender ( $P = 0.009$ ), VMS location (in the case of 400 m versus 0 m:  $P = 0.009$ ), and SV ( $P = 0.014$ ) are significantly associated with route choice behavior (turn decision versus keeping-straight decision). The odds ratio estimation of age is 0.924, which means that as the driver age increases, the probability of making turn a decision decreases. In other words, the older drivers are less willing to change driving route and the VMS guidance is less effective for them. This conclusion is consistent with the previous finding [28] but is opposed to another research result, which showed that older road users are most concerned about VMS [10]. Compared to female drivers, the male drivers have a higher turn decision rate. Therefore, the male drivers are more likely to be influenced by VMS, which is also consistent with the previous finding [28].

In the second part for VMS-II analysis, the regression results indicate that the independent variables of age ( $P = 0.007$ ), vocation ( $P = 0.012$ ), VMS information format (in the case of text-only versus text and graphics:  $P = 0.046$ ), and SV ( $P < 0.001$ ) are significantly associated with route choice behavior (turn decision versus keeping-straight decision). The age effect of VMS-II is similar to that of VMS-I. The odds ratio estimation of vocation is 7.896. It means that the professional drivers have a much higher turn decision rate than unprofessional drivers, indicating that the professional drivers are more likely to accept the information released by VMS.

According to the two logistic regression models, VMS location, VMS information formats, and speed under VMS are other three important factors that influence drivers' route choice decision. In order to illustrate how VMS location and VMS information format impact the route choice behavior, Figure 4 shows the probability of drivers' making turn decision in the case of 30-year female unprofessional drivers. It depicts that the larger speed under VMS is the less possibility that drivers choose the turn decisions. One possible explanation is that when driving too fast, drivers pay more attention to the forward road surroundings than the information released by VMSs. The other possible explanation is that if the approaching speed to intersection is too high, drivers may take more risks for changing lane to make turns at the intersection. According to Figure 4(a), it is apparent that there is a gap between 0 and 200 meters curves and 400 meters curve. It indicates that the farther distance between VMS-I and intersection, the less possibility

that drivers make turn decisions. However, when the distance between VMS-I and intersection is less than 200 m, the effect of VMS-I location will be insignificant for route choice. In Figure 4(b), when VMS information format is graphics-only, drivers are more willing to make turn decisions while drivers are not sensitive to the word message. As long as the VMS information format includes graphics road network with real-time traffic conditions, whether the VMS contains the suggested text information is not significant.

**3.2. Speed Control Behavior Analysis.** In terms of speed control behavior, this analysis focuses on the driving speed around VMS-I. The mean of the speed under VMS is 67.59 km/h, and the standard deviation is 13.12. In the cases that VMS-I's location is 0 m and 200 m from the intersection, the drivers' speed control behavior would be influenced by the intersection because they often decelerate to negotiate with the slow downstream traffic. This may confuse the VMS's effect on speed control. Therefore, the scenario C (VMS-I's location is 400 m from the intersection) is used for the analysis of speed behavior in the VMS environment. The average speed curve before and after VMS-I is shown in Figure 5. It illustrate that when drivers were approaching to VMS-I, the average speed at the beginning were increasing; however once the drivers reached 70 meters away from the VMS, the average speed had a slightly downward trend, which reflects the VMS influence on the drivers' speed control behavior. The speed curve indicates that on average, drivers perceive the VMS information and make their route choice decisions at 70 m upstream of VMS location.

**3.3. Lane Changing Behavior Analysis.** The measures of LCT, LCP, LCL, LCS, and LCD are used for exploring how VMS position and information format affect drivers' lane changing behaviors in different scenarios. The basic statistical descriptions for LCT, LCP, LCL, LCS, and LCD are summarized in Table 4. In addition, the multivariate analysis of variance (MANOVA) was conducted to analyze the effect of VMS location and VMS information format on the lane changing behavior, as shown in Table 5.

For the VMS-I, only LCD is significantly influenced by VMS location ( $F = 8.249$ ,  $P < 0.01$ ). As shown in Figure 6, when the VMS is located at the intersection (0 m), the lane changing deceleration rate is obviously higher than the other two situations ( $M = 1.09$ , S.D. = 0.7 versus  $M = 0.53$ , S.D. = 0.39 and  $M = 0.51$ , S.D. = 0.37). However, there is no obvious difference in LCD between the locations of 200 meters and 400 meters. The result indicate that if the VMS is located close to intersection, drivers perceive VMS information and make changing route decision too late so that drivers have no sufficient deceleration distance to the intersection, leading to more abrupt lane changing behaviors. Therefore, it is not appropriate to position VMS too close to diverging points in terms of traffic safety.

For the VMS-II, it is found that LCT ( $P = 0.065$ , marginally significant), LCP ( $P = 0.001$ ), LCL ( $P = 0.003$ ), and LCS ( $P = 0.001$ ) are significantly influenced by the VMS information format. However, it has no significant effect on

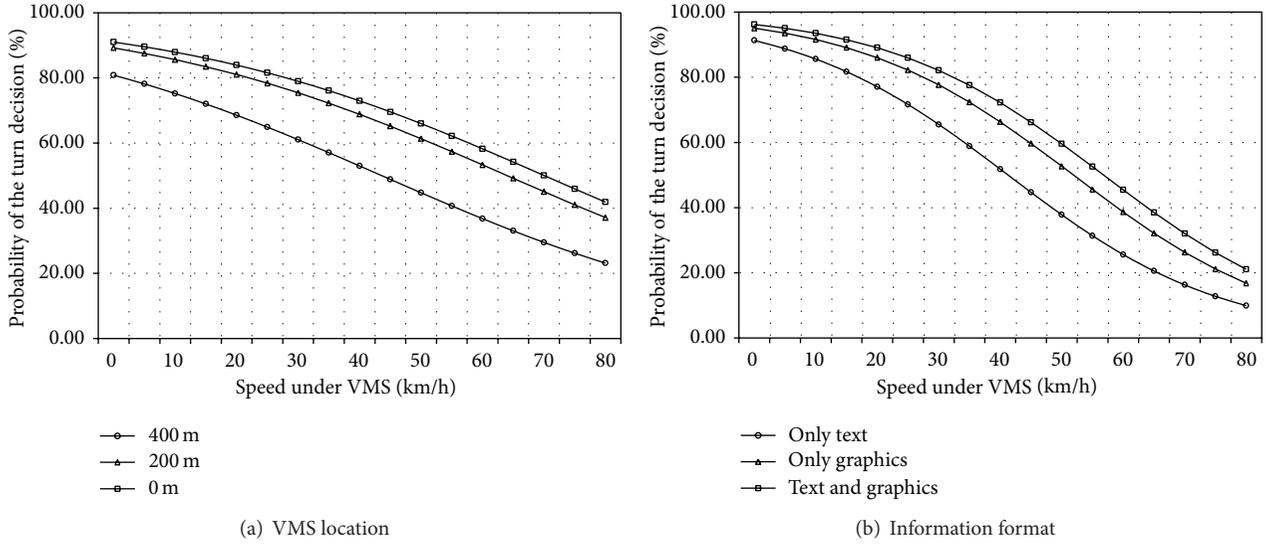


FIGURE 4: Probability of turn decision based on the logistic regression models.

TABLE 3: Parameter estimates of logistic regression models for route choice.

Mode	Variable	Level	B	S.E.	Wald	Df	Sig.	Exp(B)
The first part (VMS-I)	Age	Continuous	-0.079	0.020	16.212	1	0.000	0.924
	Gender	Male versus female	0.968	0.370	6.836	1	0.009	2.633
	Location	—	—	—	4.347	2	0.114	—
		400 m versus 0 m	-0.873	0.439	3.953	1	0.047	0.418
		200 m versus 0 m	-0.200	0.430	0.216	1	0.642	0.819
	SV	Continuous	-0.033	0.013	6.065	1	0.014	0.968
	Constant		4.682	1.341	12.197	1	0.000	107.99
The second part (VMS-II)	Age	Continuous	-0.112	0.042	7.272	1	0.007	0.894
	Vocation	Yes versus no	2.066	0.820	6.351	1	0.012	7.896
	Format	—	—	—	4.132	2	0.127	—
		Text-only versus Text and graphic	-0.885	0.442	3.998	1	0.046	0.413
		Graphic versus Text and graphic	-0.279	0.425	0.431	1	0.511	0.756
	SV	Continuous	-0.057	0.016	12.147	1	0.000	0.944
	Constant		6.597	1.775	13.816	1	0.000	733.140

TABLE 4: Descriptive statistical results for LCT, LCP, LCL, LCS, and LCD.

Variable	LCT		LCP		LCL		LCS		LCD	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Location										
400 m	12.3	4.0	246.9	187.8	151.5	94.5	12.1	6.9	0.5	0.4
200 m	14.1	6.5	244.5	155.6	160.2	101.5	11.5	5.2	0.5	0.4
0 m	13.2	5.7	179.0	65.8	145.2	65.9	11.2	3.6	1.1	0.7
Format										
Text	10.3	4.2	68.7	50.6	62.5	44.9	6.0	2.8	0.9	0.6
Graphics	14.3	5.4	158.5	99.1	129.7	78.5	9.6	5.5	0.7	0.3
Text and graphics	11.7	5.1	179.2	99.4	136.2	67.0	12.0	4.8	0.9	0.7

TABLE 5: MANOVA variance analysis of lane changing behavior.

Dependent variable	Independent variable	Type III SS	DF	Mean square	F	Sig.
The first part: VMS location (VMS-I)	LCT	38.32	2	19.16	0.625	0.539
	LCP	50245	2	25212	1.085	0.345
	LCL	2264.30	2	1132.15	0.134	0.875
	LCS	7.92	2	3.96	0.131	0.878
	LCD	3.74	2	1.87	8.249	0.001
The second part: VMS information formats (VMS-II)	LCT	142.03	2	71.01	2.890	0.065
	LCP	119829	2	59915	7.733	0.001
	LCL	57245.30	2	28622.69	6.622	0.003
	LCS	322.44	2	161.22	7.762	0.001
	LCD	0.38	2	0.19	0.602	0.551

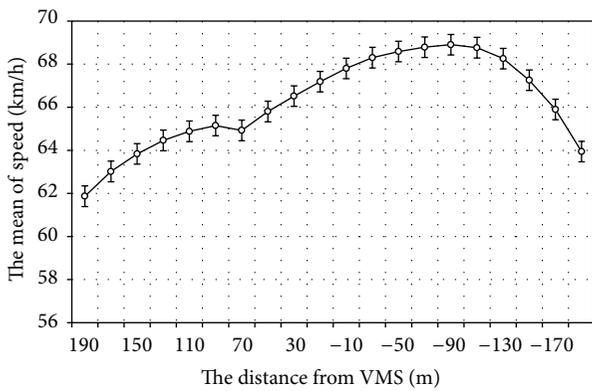


FIGURE 5: The trend of average velocities before and after VMS-I.

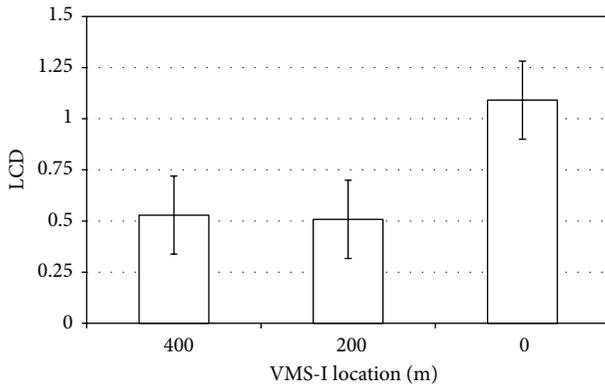


FIGURE 6: The influence of the VMS location on LCD.

LCD. The influences of VMS information format on LCT, LCP, LCL, and LCS are shown in Figure 7. It illustrates that LCT for in graphic-only format is longer than those VMS with text messages. The presumable reason is that drivers need less time to make turn decisions after perceiving the graphic information and correspondingly they can spend long time in changing lanes (see Figure 7(a)). Figures 7(b) and 7(c) show that LCP and LCL for the graphics-only and graphics-text combination formats are obviously higher than those for the text-only format, while there are no significant

differences in LCP and LCL between graphics-only and graphics-text combination formats. This finding indicates that compared to VMSs with graphic information, driver need longer reaction time and decision time to make turn decisions in the text-only VMS environment so that they change lanes closer to the intersection. Figure 7(d) shows that LCS for the text-only format is the lowest while LCS for the graphics-text combination format is the highest among the three scenarios. This finding is consistent with the trend of LCP in Figure 7(b): the farther the drivers change lane from the intersection, the higher the lane changing speed is.

#### 4. Conclusion and Discussion

The main purpose of this study is to investigate how VMS position and information format affect route choice, speed control, and lane changing behaviors in the road network using a high fidelity driving simulator. Three levels of distances ranging from 0 to 400 m between VMS location and route-diverging intersection were designed and three kinds of information formats including text-only, graphics-only, and combination of text and graphics were tested in the simulation experiment. The experimental results showed that both VMS location and information format profoundly influence driving behaviors.

In the past two decades, various experiments applied driving simulators to study the impact of VMS on the route choice behavior. Based on the experimental data, some route choice models were developed to evaluate the drivers' route choices under VMS and enhance network performance [29, 30]. Nevertheless, the previous studies mainly emphasized the effect of VMS on traffic operation but ignored the traffic safety. This study tried to explore the impacts of VMS on both traffic efficiency and safety. It was found if VMS is set too far away from route-diverging point, drivers are less likely to follow the VMS information to change driving routes. However, within the 200 m distance from route-diverging point, the effect of VMS location will be insignificant for route choice. On the other hand, if the VMS location is too close to route-diverging point, drivers would make changing route decision too late and have no sufficient deceleration distance to the diverging point. Therefore, it is suggested to position the VMS between 150 m and 200 m upstream of the

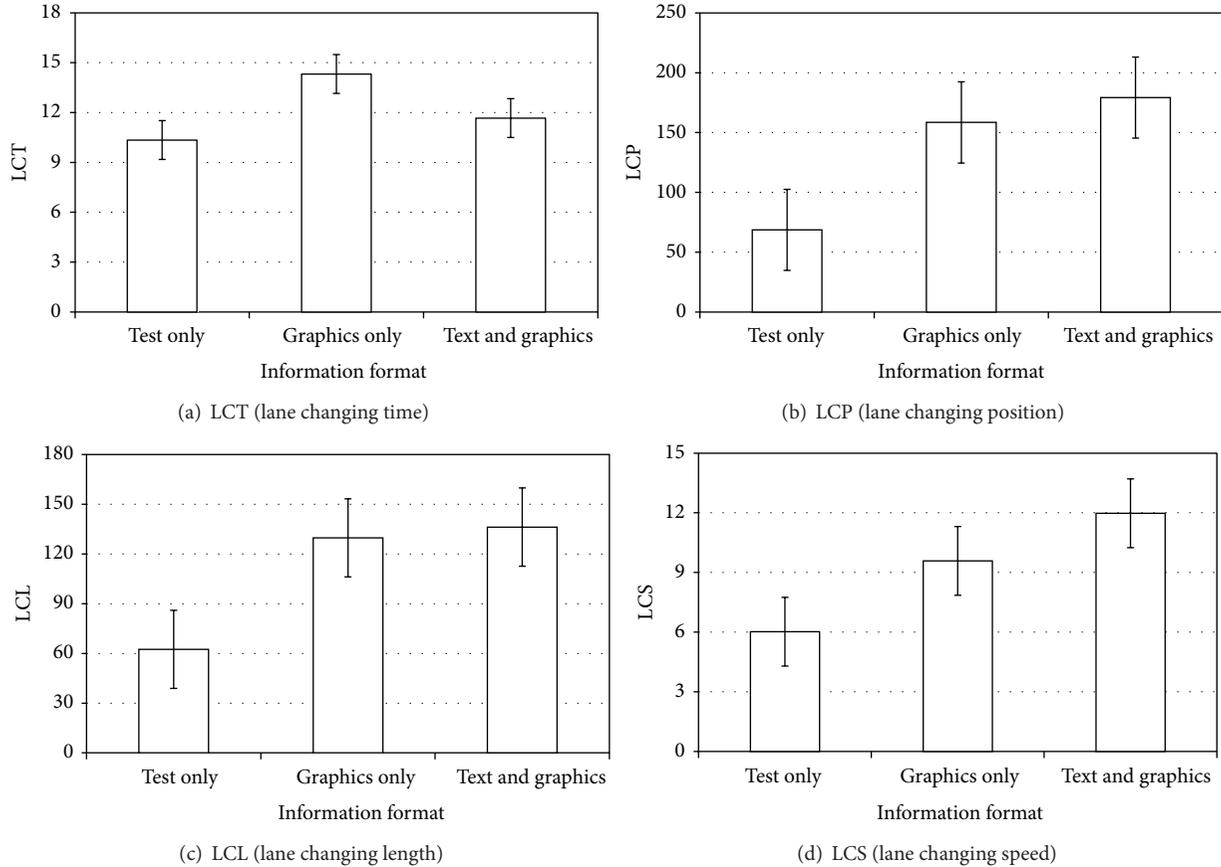


FIGURE 7: The influence of the VMS information formats on LCT, LCP, LCS, and LCA.

diverging point to balance the traffic safety and operation. Furthermore, the analysis of average speed change showed that when approaching to VMS, drivers perceive the VMS information and make their route choice decisions at a 70 m distance upstream of VMS location.

A prior study focused on how the content of VMS affected the driving behavior, which indicated that the VMS information content, including the level of detail of relevant information, socioeconomic characteristics, network spatial knowledge, and confidence in the displayed information, significantly affected drivers' willingness to divert [31]. However, all of the terms were associated with only information content, but not the information display format. An interesting finding in this paper is that drivers are more willing to changing routes with the VMS guidance information delivered by graphics than the text-only format. Also, the text-only VMS caused drivers to change lanes later and have more abrupt deceleration behaviors during the lane changing process, compared to the VMSs with graphics information. Thus, it is better to adopt the VMS design with graphics guidance information rather than the text messages only.

Additionally, the result analyses indicated that the driver characteristics of age, gender, and vocation also have significant effects on route choice behavior in the VMS environments. It was found that the older drivers are less willing to change driving route under the VMS guidance; compared to

female, the male drivers are more likely to be influenced by VMS; and the professional drivers are more likely to accept the information released by VMS.

In summary, this paper explored the relationship between driving behaviors and VMS's position and information format based on the driving simulator experiment. The findings of this study would be helpful for traffic engineers to select VMS installation locations and design VMS information delivery formats in order to optimize traffic safety and efficiency in urban road networks.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgment

This work is financially supported by "863" Research Project (2011AA110303).

### References

- [1] G. He, *ITS System Engineering Introduction*, Chinese Railway Press, Beijing, China, 2004.

- [2] H. Huang, "Dynamic modeling of urban transportation networks and analysis of its travel behaviors," *Chinese Journal of Management*, vol. 2, pp. 18–22, 2005.
- [3] R. H. M. Emmerink, P. Nijkamp, P. Rietveld, and J. N. Van Ommen, "Variable message signs and radio traffic information: an integrated empirical analysis of drivers' route choice behaviour," *Transportation Research A: Policy and Practice*, vol. 30, no. 2, pp. 135–153, 1996.
- [4] H. Shang, H. Huang, and Z. Gao, "Design real-time traffic information by cell transmission model," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 34, no. 2, pp. 234–238, 2008.
- [5] J. Jiang and L. Dong, "Characteristics and optimization of traffic flow in a two-off-ramp system," *Acta Physics Sinica*, vol. 61, no. 21, Article ID 214503, 2012.
- [6] S. Zhong, H. Ma, L. Zhou, X. Wang, S. Ma, and N. Jia, "Effects of different factors on drivers' guidance compliance behaviors under road condition information shown on VMS," *Transportation Research A: Policy and Practice*, vol. 46, no. 9, pp. 1490–1505, 2012.
- [7] P. W. Bonsall and M. Joint, "Driver compliance with route guidance advice. The evidence and its implications," in *Proceedings of the Vehicle Navigation & Information Systems Conference*, vol. 2, pp. 47–59, October 1991.
- [8] Z. Yuanfeng and W. Jianping, "The research on drivers' route choice behavior in the presence of dynamic traffic information," in *Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC '06)*, pp. 1304–1309, September 2006.
- [9] J. L. Adler, "Investigating the learning effects of route guidance and traffic advisories on route choice behavior," *Transportation Research C: Emerging Technologies*, vol. 9, no. 1, pp. 1–14, 2000.
- [10] C. Yin and X. Song, "The influence of road users' personal attributes on the cognition of VMS," in *Proceedings of the 1st International Conference on Transportation Information and Safety: Multimodal Approach to Sustained Transportation System Development-Information, Technology, Implementation (ICTIS '11)*, pp. 262–267, July 2011.
- [11] J. Chen, P. Liu, and W. Wang, "Methods for en route parking guidance and information system survey in Nanjing," *Urban Transport of China*, vol. 4, pp. 79–83, 2006.
- [12] Y. Mo and K. Yan, "Analysis of the performance of the PGIS in metropolitan area," *Road Traffic and Safety*, vol. 5, pp. 33–36, 2007.
- [13] A. Richards and M. McDonald, "Questionnaire surveys to evaluate user response to variable message signs in an urban network," *IET Intelligent Transport Systems*, vol. 1, no. 3, pp. 177–185, 2007.
- [14] W. Chen and P. J. Paul, "Analysis of a driver en-route guidance compliance and driver learning with ATIS using a travel simulation experiment," Research Report UCD-ITS-RR-97-12, Institute of Transportation Studies, University of California, Davis, Calif, USA, 1997.
- [15] B. K. Lai and W. Wong, "SP approach toward driver comprehension of message formats on VMS," *Journal of Transportation Engineering*, vol. 126, no. 3, pp. 221–227, 2000.
- [16] R. W. Allen, A. C. Stein, T. J. Rosenthal, D. Ziedman, J. F. Torres, and A. Halati, "Human factors simulation investigation of driver route diversion and alternate route selection using in-vehicle navigation systems," in *Proceedings of the Vehicle Navigation & Information Systems Conference*, vol. 2, pp. 9–26, October 1991.
- [17] R. Srinivasan and P. P. Jovanis, "Evaluation of the attentional demand of selected visual route guidance systems," in *Proceedings of the 6th 1995 Vehicle Navigation and Information Systems Conference*, pp. 140–146, August 1995.
- [18] J. Creaser, M. Rakauskas, N. Ward, and J. Laberge, "A simulator-based evaluation of smart infrastructure concepts of Intersection Decision Support (IDS) for rural thru-STOP scenarios," Project Number MN/RC-2007-31, Minnesota Department of Transportation, Duluth, Minn, USA, 2007.
- [19] V. G. B. Kolisetty, T. Iryo, Y. Asakura, and K. Kuroda, "Effect of variable message signs on driver speed behavior on a section of expressway under adverse fog conditions—a driving simulator approach," *Journal of Advanced Transportation*, vol. 40, no. 1, pp. 47–74, 2006.
- [20] S. L. Jamson, F. N. Tate, and A. H. Jamson, "Evaluating the effects of bilingual traffic signs on driver performance and safety," *Ergonomics*, vol. 48, no. 15, pp. 1734–1748, 2005.
- [21] C. Lee and M. Abdel-Aty, "Testing effects of warning messages and variable speed limits on driver behavior using driving simulator," *Transportation Research Record*, no. 2069, pp. 55–64, 2008.
- [22] A. Erke, F. Sagberg, and R. Hagman, "Effects of route guidance variable message signs (VMS) on driver behaviour," *Transportation Research F: Traffic Psychology and Behaviour*, vol. 10, no. 6, pp. 447–457, 2007.
- [23] C. Arien, E. M. M. Jongen, K. Brijs, T. Brijs, S. Daniels, and G. Wets, "A simulator study on the impact of traffic calming measures in urban areas on driving behavior and workload," *Accident Analysis and Prevention*, vol. 61, pp. 43–53, 2013.
- [24] L.-S. Tey, G. Wallis, S. Cloete, L. Ferreira, and Sicong, "Evaluating driver behavior toward innovative warning devices at railway level crossings using a driving simulator," *Journal of Transportation Safety and Security*, vol. 5, pp. 118–130, 2013.
- [25] Y. Ş. Murat and N. Uludağ, "Modeling route choice behaviour in transportation networks by using fuzzy logic and logistic regression techniques," *Technical Journal of Turkish Chamber of Civil Engineers*, vol. 19, no. 2, pp. 4363–4379, 2008.
- [26] K. Chatterjii, N. B. Hounsell, P. E. Firmin, and P. W. Bonsall, "Driver response to variable message sign information in London," *Transportation Research C: Emerging Technologies*, vol. 10, no. 2, pp. 149–169, 2002.
- [27] C. Lee, B. Ran, F. Yang, and W. Loh, "A hybrid tree approach to modeling alternate route choice behavior with online information," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 14, no. 4, pp. 209–219, 2010.
- [28] M. Yun and S. Tang, "Route diversion probability model based on guidance utility and its application," in *Proceedings of the 7th International Conference of Chinese Transportation Professionals Congress 2007: Plan, Build, and Manage Transportation Infrastructures in China*, pp. 627–636, May 2007.
- [29] H. Dia, "An agent-based approach to modelling driver route choice behaviour under the influence of real-time information," *Transportation Research C: Emerging Technologies*, vol. 10, no. 5–6, pp. 331–349, 2002.
- [30] K. V. Katsikopoulos, Y. Duse-Anthony, D. L. Fisher, and S. A. Duffy, "The framing of drivers' route choices when travel time information is provided under varying degrees of cognitive load," *Human Factors*, vol. 42, no. 3, pp. 470–481, 2000.
- [31] S. Peeta, J. L. Ramos, and R. Pasupathy, "Content of variable message signs and on-line driver behavior," *Transportation Research Record*, no. 1725, pp. 102–108, 2000.

## Research Article

# Effect of Height on Pedestrian Route Choice between Stairs and Escalator

Qian Li,<sup>1</sup> Changxu Ji,<sup>2</sup> Limin Jia,<sup>1,2</sup> and Yong Qin<sup>1</sup>

<sup>1</sup> School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup> State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Limin Jia; jialm@vip.sina.com

Received 18 November 2013; Revised 22 January 2014; Accepted 10 February 2014; Published 13 March 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Qian Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to overcome the subjectivity of existing pedestrian route choice models, an alternative choice model is presented based on the utility equation. It is composed of several indirectly objective characteristic variables, including the height, length, and width of interlayer facilities; speed of automated facilities; and carry-on luggage. Considering the scene that pedestrians choose between the stairs or escalators, an extended binary logit model is developed. Calibration and validation of the model are accomplished by using the data collected in four typical passenger transfer stations in Beijing, China. The results show that the proposed model has an average accuracy of 86.56% in bidirection for predicting pedestrians' behavior. An interesting phenomenon can be found that the length of facility has poorer impact than height on pedestrians' route choice behavior. Some quantitative and irradiative conclusions have been illustrated on the relationship between the selection probability and the variables, which is expected to be valuable for extracting the implicit theoretical mechanism of passenger choice behavior.

## 1. Introduction

Integrated development of the modern passenger transfer hubs realizes the combination of intercity and in city transportation within or in the same space layout. Centralized transferring of pedestrians greatly improve the operating efficiency of the urban passenger transport network. Passengers transfer space has a multilayer and attaches a certain amount of interlayer facilities connection, such as stairs, escalators, elevators and slopes. The three-dimensional structure increases the complexity of the traveler's behavior and activities in the hub. At peak time, due to the impact of a large number of passengers, bottlenecks usually appear at interlayer facilities, which induces local congestions and unnecessary delays. Therefore, understanding of the role of the relationship between these facilities and pedestrian behavior has become a very compelling issue [1–4], as the dynamic interactions of person-to-person [5, 6] and vehicle-to-vehicle [7]. Although these problems are very specific, they also implied considerable theoretical mechanism. Mastering this knowledge can help building designers to design more

reasonable structure of facilities, make pedestrians use the facilities relatively balanced, promote the improvement of critical facilities' level of service [4], and finally let pedestrians get enjoyable experience in the process of travel. In order to achieve this goal, it becomes very important to define the characteristics of passengers' choice behavior among interlayer facilities, and to estimate the effect of physical properties of interlayer facilities on the choice behavior.

As is known to all, in the process of pedestrians through level bridging facilities, the physical exertion must be paid by the tourists that have high-positive correlation with the height of facilities, which plays a key role in their choice decisions, especially those carrying luggage. Moreover, this must be related to the fundamental structure of facilities and current environmental factors such as flow density. There are lots of researches on modeling and analyzing the passenger choice behavior of facilities. Daamen et al. [3] built utility analysis model for selection problem of the entire path of the longitudinal facilities. Seneviratne and Morrall [8] and Guo et al. [9] analyzed detailed influencing factors of pedestrian route choice behavior. Borgers and

Timmermans [10] established the selection model of shortest walking distance in horizontal dimension. Cheung and Lam [11] took pedestrians in Hongkong as an example, studied the passenger facilities choice behavior, and proposed the perceive time as an influence factor. Ye et al. [12] analyzed the traffic characteristics of the passenger of stairs and corridors in the Shanghai subway station and established the relationship between the interlayer facilities and the volume of passenger traffic. Through researching the relationship between the interlayer facilities layout and passengers' route choice in the transfer station, such as stairs and escalators, Daamen [1] and Zeiler et al. [13] summed up the degree of passenger congestion of interlayer facilities as a key factor directly impact on behavior of passengers. Jia et al. [14] based on comprehensive analysis characteristics of the passenger flow and service status of the inner facilities in the hub established a systematic analysis method of capacity adaptability of facilities' ability. Lam and Chau [15] according to the variation of pedestrian flow of shopping center in station developed multiple walking facilities control strategy to reduce resource waste. In addition, Olander and Eves [16] and Lewis and Eves [17] analyzed the influence of choice-on-point prompt on enhancing the use of stairs by pedestrians, from the perspective of human health.

In short, these studies have focused on the relationship between passenger flow and interlayer facilities to reveal the theoretical mechanism of passenger choice behavior, ignoring the effect of interlayer facilities physical properties on passengers. As for the models, discrete utility theory has been extensively used on these issues, whether in horizontal or vertical direction. The difference between these models is the selection of influencing factors on calculation utility function, such as the generalized consumption [3], shortest walking distance [10], and generalization of time [11]. All of these did not positively consider the height of facilities which is a quantitatively specific parameter. Take the perception time, for example, since it is a highly technical matter for people to reasonably map the objective existence of physical exertion paid by climbing stairs to the subjective perception time, even if the professional architect also hardly completes the transform, the results of theoretical analysis in the application of engineering practice are at stake.

On the basis of previous research, the height of facilities was proposed as an independent factor in a utility model to describe the relationship between height of facilities and pedestrian facilities choice behavior.

After a brief literature review, the new model was put forward, and some of the features were expounded in detail. Then, the scene of the traffic data which model calibration required was introduced, and the model calibration results were shown. Finally, theoretical characteristics were illustrated which extracted from further analysis of the model and came to some meaningful conclusions.

## 2. Model and Calibration

*2.1. Analysis of Influence Factors.* To calculate the utility function, we need select the influencing factors. These is the most fundamental problem. An extended literature study on

empirical data and modeling of pedestrian route choice can be found in [1, 3, 8]. Generally, three sets of characteristics would be taken into concern, including the facility structure, basic consumption, and passenger property.

*(i) Facility Structure.* The number of routes, one of the most important factors, is related to the categories of the mathematical model. Length of facility is closely related to the time which passenger need to get through the facility, also it is a primary factor to analyze the choice behavior and more important in the pedestrian route choice behavior model in horizontal dimension [10].

*(ii) Basic Consumption.* Time consumption at least consists of two parts, including the travel time and waiting time, which are under the influence of the length of facilities, passenger current speed, channel capacity, and the number of waiting for lineup.

Passengers will take the corresponding physical strength when they are using level bridging facilities. Within reasonable delay time, people tend to choose less expensive physical facilities. That is to say, physical consumption plays a substantial role in pedestrians' choice in vertical dimension. If passengers carrying luggage (we consider only the heavy luggage without light handbag), more energy will be consumed when passengers go through the non-automated facilities.

*(iii) Passenger Attribute* [3, 8]. The effects of differences in gender and age of passengers are not obvious. Also, trip purpose plays no role for work-related walking trips. This paper focuses on the ordinary pedestrians without difference of purpose.

After an overview of the influencing factors above, the structure and basic consumption are selected for subsequent modeling; passenger property will not be considered. Even, the items of data collection are corresponding with them.

*2.2. Binary Logit Model.* Theory of disaggregate model is based on the hypothesis that consumers choose to pursue utility maximization. In the issue of the passenger facilities selection, passenger choice behavior and consumer have the same principle. Logit model, which has advantages of simple structure and strong applicability, divides utility into uncertainty utility and random utility and assumes that the random utility obeys certain probability distribution, thus obtains the probability of travelers choosing different transport facilities.

Assume that the passengers can select the options independently, the utility model  $U_{in}$  of the passenger  $n$  selection facilities  $i$  can be expressed as

$$U_{in} = V_{in} + \varepsilon_{in}, \quad (1)$$

where  $V_{in}$  is the fixed term in utility function of traveler  $n$  that selects facility  $i$  and  $\varepsilon_{in}$  is the random term in utility function of traveler  $n$  that selects facility  $i$ .

Constructing and evaluating the utility function is a key link in the process of correctly analyzing passenger choice behavior. Borgers and Timmermans [10] formulate a model for pedestrian route choice in city centers given pedestrian's

TABLE 1: Characteristic variables.

Characteristic variables	Variables represent	
	Escalator	Stairs
Time consumption		
Walking time ( $t_1$ )	$t_1^{\text{esc}}$	$t_1^{\text{st}}$
Delay time ( $t_2$ )	$t_2^{\text{esc}}$	$t_2^{\text{st}}$
Physical consumption		
Height ( $h$ )	0	$h^{\text{st}}$
Luggage ( $g$ )	0	$g^{\text{st}}$

destination. It is assumed that route choice is primarily influenced by the distance of the alternative routes. Cheung and Lam [11] adopt a logit model to model pedestrian's choice between escalators and stairs in a station, in which the utility of an alternative  $U_r$  is described by

$$\begin{aligned}
 U_r &= V_r + \varepsilon_r \\
 \text{with, } V_r &= \alpha + \beta T_r \quad (2) \\
 \text{with, } T_r &= t_{r,\text{delay}} + t_{r,\text{walk}},
 \end{aligned}$$

where  $\alpha$  and  $\beta$  are parameters to be estimated,  $T_r$  is the travel time along route  $r$ ,  $t_{r,\text{delay}}$  is delay on the walkway leading to escalator or stair,  $t_{r,\text{walk}}$  is the travel time along escalator or stair. Rewriting the resulting logit model leads to the following expressions:

$$\begin{aligned}
 P_{\text{st}} &= \frac{\exp(V_{\text{st}})}{\exp(V_{\text{st}}) + \exp(V_{\text{esc}})} = \frac{1}{1 + \exp(V_{\text{esc}} - V_{\text{st}})} \quad (3) \\
 P_{\text{esc}} &= 1 - P_{\text{st}},
 \end{aligned}$$

where  $P_{\text{st}}$  is the probability of pedestrian choosing escalator,  $P_{\text{esc}}$  is the probability of pedestrian choosing stairs, and  $(V_{\text{esc}} - V_{\text{st}})$  is relative discomfort measure for using stairs.

Combining with the characteristics of passenger choice behavior in vertical dimension, the time and physical consumption are select and used as characteristic variables, as shown in Table 1. Time consumption consists of walking time  $t_1$  and delay time  $t_2$ . Physical consumption consists of interlayer height  $h$  and carry-on luggage  $g$ . In this work, passengers within automated facilities are regard as having the physical consumption as zero, and only distinguish whether passengers carrying luggage or not. 0 for nonluggage, otherwise is 1.

Then, the new  $V_r$  is replaced with

$$V_r = \theta_0 + \theta_1 (t_1 + t_2) + \theta_2 (h + g), \quad (4)$$

where  $\theta_0$ ,  $\theta_1$ , and  $\theta_2$  are parameters to be estimated.

Set

$$\begin{aligned}
 V_{\text{esc}} - V_{\text{st}} &= \theta_0 + \theta_1 \Delta T + \theta_2 \Delta P \\
 \Delta T &= \{(t_1^{\text{esc}} + t_2^{\text{esc}}) - (t_1^{\text{st}} + t_2^{\text{st}})\} \\
 &= \{(t_1^{\text{esc}} - t_1^{\text{st}}) + (t_2^{\text{esc}} - t_2^{\text{st}})\} \\
 &= \Delta t_1 + \Delta t_2 \\
 \Delta P &= \{(h^{\text{esc}} + g^{\text{esc}}) - (h^{\text{st}} + g^{\text{st}})\} \\
 &= -(h^{\text{st}} + g^{\text{st}}).
 \end{aligned} \quad (5)$$

Substituting (5) into (3), (6) will be available as

$$\begin{aligned}
 P_{\text{esc}} &= \frac{1}{1 + \exp(\theta_0 + \theta_1 \Delta T_1 + \theta_2 \Delta T_2 + \theta_3 h^{\text{st}} + \theta_4 g^{\text{st}})} \quad (6) \\
 P_{\text{st}} &= 1 - P_{\text{esc}}.
 \end{aligned}$$

### 2.3. Model Estimation

**2.3.1. Basic Data.** To exclude the interference of other transportation modes and obtain pedestrian traffic flow parameters in a range as wide as possible, four transfer stations were selected as the observation sites in Beijing, such as Xizhimen subway station, Beijing South Railway Station, Zoo Station, and Zhichun Road Station, for they have large amounts of vertical pedestrian travel demand. And, the number of different choice situations and route alternatives is reasonably high in these stations.

The way of basic data acquisition is video observations, as described by Cheung and lam [11]. The pedestrian flows were recorded with Digital Camera for an hour, respectively, during the peak times and the information in the video were extracted manually and organized in the form of individual cases afterward. 3849 passengers' choice decisions, 2207 in ascending and 1642 in descending, had been extracted from 40 video files of 10 scenes allocated in the four stations. Some snapshots can be seen in Figure 1. The data has been split into two groups, one is for calibrating and the other for validating.

The factors considered as well as the corresponding data obtained from the videos in the study include date, peak time, station, direction, structure of facilities (height, length, and width), speed of automated facilities, the number of people in queue for escalator, and whether carry-on luggage.

**2.3.2. Model Calibration.** Model calibration is divided into three steps. Firstly, the form of the utility function and the characteristic variable must be determined, on behalf of calibration of the model. Secondly, using the maximum likelihood estimation method to calibrate the parameters and making some preparations for model validations. Finally, the covariance matrix is used to make  $t$  test or other tests. The results of model test are shown in Table 2.

As the absolute value of the  $t$ -test is larger than 1.96, there is 95 percentage degree of confidence on indicating that the corresponding variable is a factor that can affect the probability of selection. In Table 2, since the absolute values of

TABLE 2: Calibration result.

Direction	Type	$\theta_0$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
Ascending	Value	6.6324	-0.2501	-0.5986	0.8642	0.9976
	$t$ -test	2.16	0.51	2.53	2.86	-3.76
	Accuracy	89.41%	Goodness	0.86	Conformity	0.81
Descending	Value	5.9077	-0.5539	-0.7086	0.7331	0.8244
	$t$ -test	2.26	-0.84	-4.15	2.17	3.93
	Accuracy	83.70%	Goodness	0.79	Conformity	0.76



(a) In ascending



(b) In descending

FIGURE 1: Video snapshots.

$t$ -test value on parameter  $\theta_1$  are less than 1.96 in bidirections, the walking time  $t_1$  is wiped off from the characteristic variables for its little affect. In terms of accuracy, the average accuracy of the models for forecasting the passengers' choice decisions is 86.56%, which meets the precision requirements of BL model. The results of model calibration are as follows.

In ascending direction

$$P_{st} = \frac{1}{1 + \exp(6.6324 - 0.5986\Delta t_2 + 0.8642h + 0.9976g)}$$

$$P_{esc} = 1 - P_{st}. \quad (7)$$

In descending direction

$$P_{st} = \frac{1}{1 + \exp(5.9077 - 0.7086\Delta t_2 + 0.7331h + 0.8244g)}$$

$$P_{esc} = 1 - P_{st}. \quad (8)$$

### 3. Analysis of Choice Mechanism

This section presents an examination of various properties of pedestrian choice behavior indicated in the percentages of the stairs that will be used. In line with [3, 11] the proposed model also showed that passengers are more inclined to choose escalators as they improve experience, more obvious in upward direction. For distinctively analyzing the performance of level bridging facilities, there are three factors that would be highlighted in the following, such as height of facilities, delay time, and carry-on luggage. A specific set of parameters to a passenger will produce a variety of results, which revealed some passenger's preferences on choice behavior.

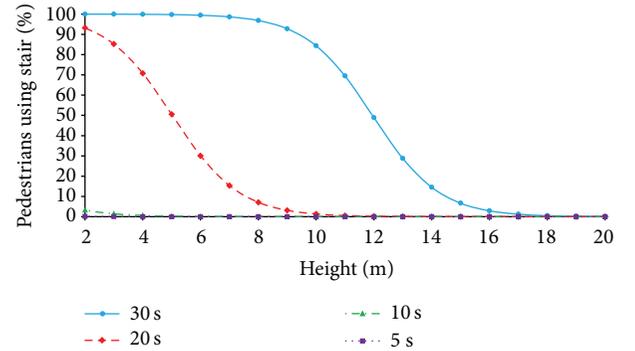


FIGURE 2: Percentage of pedestrians using stairs by different waiting time.

**3.1. Varying Facilities Height.** Figure 2 shows the percentage of pedestrians using stairs plotted against height of vertical access settings without luggage in ascending by delay time at 5, 10, 20, and 30 seconds, respectively. Basically, with the passengers' delay time in front of escalator growing, the probability that they choose the stairs becomes higher.

When height is 5 meters and delay time is 20 seconds, the probabilities of each alternative are chosen almost the same. Once the balance is broken, passengers will more clearly to make a decision. In case of ignoring waiting time, the vast majority of people will not pay attention to the stairs. Escalator plays an important role in the process of transition between the layers. As the delay time changes from 10 to 20 seconds or more, the concern of pedestrians shows a dramatic shift, so much as more than 98 percentages of people choose stairs when the waiting time is for 30 seconds. Under the condition that the floor is not too high,

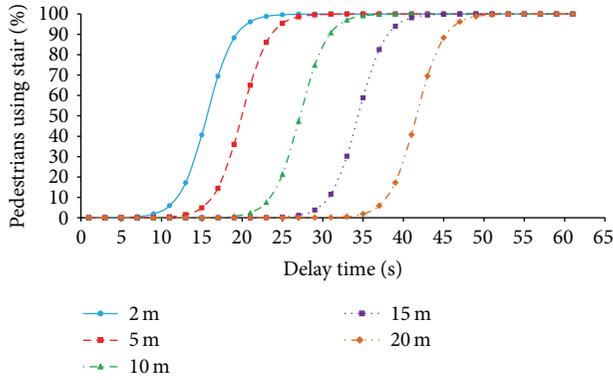


FIGURE 3: Percentage of pedestrians using stairs by different height.

it is necessary to appropriately increase the stairs width to encourage passengers to get through as soon as possible. Otherwise, a lot of passengers were forced to crowd in front of the facilities, although they are willing to pay extra strength in climbing the stairs.

**3.2. Varying Delay Time.** Figure 3 shows the percentage of pedestrians using stair plotted against delay time without luggage in ascending by height at 2, 5, 10, 15, and 20 meters, respectively. The height level bridging facility is larger, the passenger’s acceptable delay time standing by escalator is longer, and the number of passengers in front of selector is more, under the selection probability of stairs is less than 15%.

For example, as the height reaches 20 meters, the traveler’s acceptable delay time will not exceed 35 seconds. Otherwise, a significantly increase occurs in probability of passengers choosing stairs. Then, a larger number of passengers will gather in the access corridor to the escalator.

In the design stage of infrastructures, corresponding to the height of the interlayer facilities, the area of access facilities to level bridging facilities must be large enough, at least for satisfying the demands of the corresponding number of passengers.

**3.3. Carry-On Luggage or Not.** Figure 4 shows the percentage of pedestrians using stairs against height of interlayer facilities with 20 seconds delay time in ascending, including the cases that they carry-on luggage or not, respectively. In these certain conditions, the percentage of choosing stairs in passenger with load is lower than lightweight ones within a certain range. As the interlayer height is 6 meters, the 53.85 percentage of passengers without luggage will take into account stair, and only 30.09 percent of passengers with luggage want to climb the stairs to avoid the delay time. Therefore, if the height of target facilities is more than 10 meters, almost all travelers tend to abandon the use of stair rationally, no matter whether they carry luggage or not. So, when the height reached a certain value, passengers’ choice behavior will be convergence.

According to the analysis on effect of luggage to pedestrian’s choice behavior, a good proposal could be put forward.

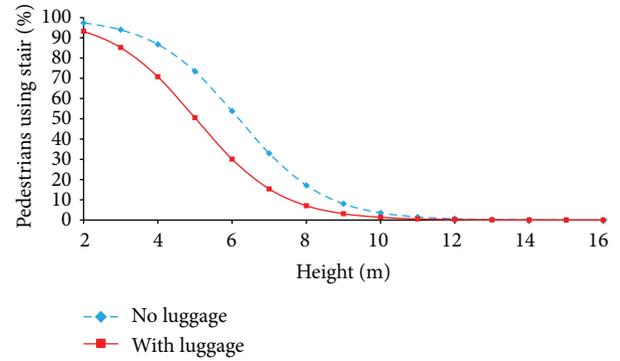


FIGURE 4: Effect of luggage on probability of passenger using stairs.

Compared to commuter staff, the proportion of intercity travelers which carrying luggage is higher and the number of escalator in intercity transportation hub should be more than transfer hubs in city, in order to improve the capacity of the automation facilities and provide better service to passengers.

#### 4. Conclusions

This study developed a binary logit model with an alternative utility function to predict pedestrian’s choice behavior between stair and escalator in public transfer stations. Different from the other models, the generalized extreme value indicated by subjective perception of time consumption of passenger uses level bridging facilities that have been divided into two parts. One is time consumption which consists of walking time and delay time, and the other is physical consumption which is represented by the height of facilities and whether carrying-on luggage. In addition, ascending and descending data for calibrating and testing the function has been collected in five passenger transfer stations in Beijing, China, including ten stairs and escalator parallel scenarios.

According to the results, the average accuracy of 86.56% is extraordinary high for this type of model. The model also shows that the length of facility has little influence on passenger selection behavior. So, it is reasonable that the height is identified as an independent characteristic variable. Some phenomena have been observed with further analysis of the model. Passengers are more sensitive to facilities height in ascending than in descending. Due to the objectivity of the selected parameter, we can intuitively understand the relationship between pedestrian choice behavior in vertical dimension and their specific environment. Also, the estimated values for infrastructure types may directly be applied in a simulation tool, even as a theoretical reference for facility design decisions.

#### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was supported by the Independent Research Topic of State Key Laboratory (RCS2011ZT002), National Nature Science Foundation of China (71171015), and National High Technology Research and Development Program (2011AA110506).

## References

- [1] W. Daamen, *Modelling Passenger Flows in Public Transport Facilities*, Delft University of Technology, Delft, The Netherlands, 2004.
- [2] W. Daamen, P. H. L. Bovy, and S. P. Hoogendoorn, "Choices between stairs, escalators and ramps in stations," in *Proceedings of the International Conference on Computer System Design and Operation in the Railway and Other Transit Systems*, pp. 3–12, Prague, Czech Republic, July 2006.
- [3] W. Daamen, P. H. L. Bovy, and S. P. Hoogendoorn, "Passenger route choice concerning level changes in railways stations," in *Proceedings of the Transportation Research Board Annual Meeting*, pp. 1–18, National Academy Press, Washington, DC, USA, 2005.
- [4] Transportation Research Board, *Highway Capacity Manual 2000*, Transportation Research Board, Washington, DC, USA, 2000.
- [5] D. Helbing, L. Buzna, A. Johansson, and T. Werner, "Self-organized pedestrian crowd dynamics: experiments, simulations, and design solutions," *Transportation Science*, vol. 39, no. 1, pp. 1–24, 2005.
- [6] V. J. Blue and J. L. Adler, "Cellular automata microsimulation for modeling bi-directional pedestrian walkways," *Transportation Research B*, vol. 35, no. 3, pp. 293–312, 2001.
- [7] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [8] P. N. Seneviratne and J. F. Morrall, "Analysis of factors affecting the choice of route of pedestrians," *Transportation Planning & Technology*, vol. 10, no. 2, pp. 147–159, 1985.
- [9] H. Guo, W. Wang, W. Guo, X. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [10] A. W. J. Borgers and H. J. P. Timmermans, "City centre entry points, store location patterns and pedestrian route choice behaviour: a microlevel simulation model," *Socio-Economic Planning Sciences*, vol. 20, no. 1, pp. 25–31, 1986.
- [11] C. Y. Cheung and W. H. K. Lam, "Pedestrian route choices between escalator and stairway in MTR stations," *Journal of Transportation Engineering*, vol. 124, no. 3, pp. 277–285, 1998.
- [12] J. H. Ye, X. Chen, C. Yang, and J. Wu, "Walking behavior and pedestrian flow characteristics for different types of walking facilities," *Transportation Research Record*, vol. 2048, no. 6, pp. 43–51, 2008.
- [13] I. Zeiler, C. Rudloff, and D. Bauer, "Modeling random taste variations on level changes in passenger route choice in a public transport station," in *Pedestrian and Evacuation Dynamics*, pp. 185–195, 2011.
- [14] H.-F. Jia, B.-F. Sun, Q.-Y. Luo, and J.-C. Han, "Measurement and adaptation analysis of their facilities capacity for the subway passenger terminal," *Journal of Jilin University*, vol. 39, no. 2, pp. 199–203, 2009.
- [15] B. Y. M. Lam and K. W. Chau, "Explaining the variations in the pedestrian flow values of shopping centres," *Facilities*, vol. 30, no. 3–4, pp. 164–176, 2012.
- [16] E. K. Olander and F. F. Eves, "Elevator availability and its impact on stair use in a workplace," *Journal of Environmental Psychology*, vol. 31, no. 2, pp. 200–206, 2011.
- [17] A. L. Lewis and F. F. Eves, "Testing the theory underlying the success of point-of-choice prompts: a multi-component stair climbing intervention," *Psychology of Sport and Exercise*, vol. 13, no. 2, pp. 126–132, 2012.

## Research Article

# Optimizing Crew Rostering with Multilicense on High-Speed Railway Lines

**Zhiqiang Tian**

*School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China*

Correspondence should be addressed to Zhiqiang Tian; tianzhiqiang\_1128@126.com

Received 18 November 2013; Accepted 1 January 2014; Published 11 March 2014

Academic Editor: Huimin Niu

Copyright © 2014 Zhiqiang Tian. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper studies the crew rostering problem with the consideration of multilicense of crewmembers. A 0-1 integer programming model is established to minimize the number of crewmembers used and to maintain the working time balance and the income balance of crewmembers. The method for calculating the reasonable cycle schemes is designed by considering the parameters of monthly working time standard and the number and the average working time of crew routes. The order for selecting the optimal cycle scheme is then determined with consideration of the connection relationships between crew routes. According to the characteristics of the problem and the requirements in application, this paper presents the design of an improved ant colony algorithm for solving the optimization model. The reasonableness of the model and the effectiveness of the algorithm are verified by a numerical example with 43 generated crew routes.

## 1. Introduction

Crew planning problem (CPP), one of the classic problems in operations research, is common in airlines, railway, urban public transport, and other transport industries. After costs for fuel, crew costs constitute the second largest expense of an airline [1]. An annual saving of tens of millions of US dollars for large airlines can be obtained by a good crew plan; hence, the research and application of the CPP in airlines start at an early stage. With the development of other transportation industries, their CPPs have received wide attention.

Due to its complexity, CPP is usually divided into *crew scheduling* and *crew rostering* phases [2, 3]. Crew scheduling plan usually combines the minimum units (e.g., flight legs and crew segments) into a task (e.g., flight duty and crew route) that can be accomplished in a day by one crewmember, whereas crew rostering plan assigns these tasks to crewmembers in a given cycle scheme (e.g., a week, a month, or a custom cycle) by satisfying various criteria. So far, many studies have focused on crew scheduling problem [2–5]. The operation cost of crewmembers can be saved greatly by a reasonable combination of work units. However, research on crew rostering problem started lately. Although

the operation cost can still be saved by controlling the number of crewmembers, the workload and life quality of the crewmembers restrict the optimization degree of the costs.

Ernst et al. [6] presented a review of staff scheduling and rostering. They pointed out that rostering algorithms must be more general in the future because of the changing working environments and conditions. In practice, each problem has its own business characteristics and core problems to be solved. Thus, the rostering method is still developed for certain problems.

According to the different uses of crewmembers, the crew rostering problem can be solved in various ways following different approaches. For example, Emden-Weinert et al. [7] introduced a rostering system for tram and bus drivers at the Bremer Straßenbahn AG in Germany. The proposed system divides the drivers and duties at a depot into groups and then develops rotas for each group. In this study, we call this process *grouped rostering*, in which a rota is a pattern of rest days and shift type. However, individual rosters at most European airlines are constructed directly for each crewmember; this process is called *personalized rostering* [8–10]. In this situation, all crewmembers should have their own unique evaluation criteria or individual needs. All individual

needs of crewmembers should be satisfied as much as possible to generate feasible rotas. Chu [9] dealt with a personnel planning problem in an international airport by decomposing it into a duty generating phase (a GP planner), followed by GP scheduling and rostering phase. Maenhout and Vanhoucke [11] indicated that the objective of this problem is to assign a personalized roster to each crewmember, which minimizes the overall operational costs while ensuring the social quality of the schedule. They also compared the proposed scatter search algorithm with optimal solutions obtained by an exact branch-and-price procedure and a steepest descent variable neighborhood search. Chu et al. [12] proposed another type of crew rostering plan with a single cycle, which combined all crew routes into a single loop. Although this method could realize the absolute equilibrium of working time, the cycle of the plan is determined by the number of crew routes, which reduces the flexibility and adjustability of the plan. The common feature of the former two approaches is that the cycle scheme of the plan is determined before planning. Therefore, for the convenience of description, the first two approaches are called crew rostering with given cycle.

Early studies on staff rostering assumed that each worker can only perform one type of task. However, with the improved crewmember capability and staff training, problems related to multiskilled staff scheduling have attracted attention. According to the difference of license types, the multilicense crew rostering problem may have the following different cases. (a) All crewmembers have only one license type, and these license types have no inheritance and substitutability. In working out the crew rostering plan with this characteristic, all crew routes will be divided into groups according to their license types and work out crew rostering plan. (b) Contrary to the first situation, all crewmembers have received full training and thus can undertake crew routes with any license type. At this time, the constraint of license type does not influence the feasibility of the plan, and all crew routes can be placed in one set with no differences when working out the crew rostering plan. (c) Each crewmember possesses part of all licenses types, and some of them have the inheritance and substitutability with other licenses. That is, crewmembers with higher license types can accomplish crew routes that need the lower one. When working out crew rostering plan with this characteristic, the different combinations of the crew routes with different license types will significantly affect the result of crew rostering plan. Corominas et al. [13] addressed a problem of assigning different task types to completely multiskilled workers in a retail chain selling clothes. Avramidis [14] addressed a multiskill staffing problem in a call center and developed a two-stage heuristic that finds good solutions to the founded mathematical programs.

From a modeling point of view, crew rostering problem is usually formulated as a set partitioning problem and its transformation. Various types of crew rostering plan have different objective functions. Buhr [15] treated this problem as a zero-one integer programming problem with a defined objective function and a specific set of constraints. Beaumont [16] used mixed integer programming to design employee rosters in the rostering plan, whereas Ryan [17] and

Kohl and Karisch [1] modeled the crew rostering problem as a generalized set partitioning problem. From a solution point of view, the algorithm for solving the crew rostering problem can be generally divided into three categories, namely, accurate solution algorithm, heuristic algorithm, and modern optimization algorithm. The accurate solution algorithm can find an optimal solution because the scale of the problem is small. However, finding an optimal solution with this method is hard to be accepted for its long solution time when the numbers of variables and constraints are very large. Thus, the heuristic algorithm [18, 19] and modern optimization algorithm [20–23] are designed to solve this problem more accurately and more quickly.

The remainder of this paper is organized as follows. A detailed statement of the problem is described in Section 2. A mathematical model of the crew rostering problem based on the analysis of the optimization targets and constraints of our problem is established in Section 3. In Section 4, two algorithms are designed to decide a reasonable cycle scheme and compile the crew rostering plan under a given cycle. Finally, a numerical example with generated data is presented in Section 5.

## 2. Problem Statement

For describing the crew rostering problem in our research, the descriptions of some concepts are given below. *Crew route* is a train line (or a sequence of train lines) for a crewmember to accomplish during a day's work, which is obtained by the combination of the working segments. *Crew roster* is a long term plan for the operation and management of crewmembers, which arranges the work of crewmembers in a fixed cycle scheme according to the related regulations of high-speed railway lines. *Rota* is a sequence that consists of several crew routes, which are finished by a group of crewmembers in a circulation with the satisfaction of working time and resting time. A rota usually contains the following information: number, working time, and income of all crew routes. *License type* is the requirement of the skills of crewmembers when accomplishing a specific crew route. Different license types represent different skills of crewmembers. The higher the level of license required by a crew route, the higher the revenue per unit time of the crewmember.

In China, the train's running time on high-speed railway lines ranges from 6:00 to 24:00, which indicates that the working time of crewmembers is also in this range. Therefore, crewmembers can accomplish two crew routes in two adjacent days with the satisfaction of resting time standard. Then, the crew rostering plan with a given cycle scheme has its natural advantage of arranging the crewmembers' daily work.

The relationships are built directly between crew routes and crewmembers in a personalized crew rostering plan. This method can minimize the number of crewmembers used. However, the scale of the problem is too large to design an effective algorithm for finding an optimal solution. The grouped crew rostering plan can largely reduce the scale of the problem. It can also help implement the plan and manage the crewmembers. However, the cycle scheme of grouped

rostering must be designed as reasonable as possible to avoid the undesirable increase in the number of crewmembers. The grouped rostering problem is studied in this paper based on the determination of the optimal cycle scheme to generate and organize the crew rostering plan more easily.

On high-speed railway lines, the Automatic Train Operation (ATO) system can replace the drivers in some running sections under the supervision of the Train Automatic Protection system. The application of ATO not only realizes the efficient and energy-saving operation of the train but also reduces the intensity of the drivers, allowing drivers to work on different trains. Under this circumstance, this paper proposes a crew scheduling problem of motor train set drivers with multilicense.

Based on the above analysis, this paper studies the crew rostering problem of motor train set drivers with the following characteristics: (a) the reasonable cycle scheme can be determined and (b) all crewmembers are fully trained with the same driver licenses and (c) used as groups in the plan.

Tian and Niu [24] studied the crew rostering problem with a given cycle. However, a slightly complicated two-step optimization approach is adopted to avoid the inactivity of some crewmembers because the cycle scheme is given in advance without considering the real data. In this research, we first determine the reasonable cycle scheme of the crew rostering plan (scilicet, the value of the planning horizon  $T$  and the work days  $T_w$  in the plan), which not only helps to reduce the number of crewmembers used but also ensures that the working time of the crewmembers satisfies the monthly standard. Then, the establishment of the crew rostering plan is carried out with the given cycle scheme. The works in the rota are accomplished by a group of crewmembers. Thus, the objective of the crew rostering plan is to find a partition of all crew routes. This partition ensures that the crew routes in each group can form a feasible rota by appropriate connections and that the working time and income are balanced with other rotas.

### 3. Optimization Model

*3.1. Optimization Model.* The number and the operation cost of staff have always been the main objectives for the staff scheduling and rostering problem. For the crew rostering problem on high-speed rail lines, the operation cost is composed of two parts: the cost for hiring the crewmembers and the cost for finishing the tasks. The fees paid to the crewmembers by the company for accomplishing the crew routes are proportional to the working time and the license type, which are constant and independent of the number of crewmembers. However, the cost for hiring the crewmembers is relevant to the number of crewmembers. The tasks in crew routes are usually different from each other, and their corresponding working times are also different. The fairness of the crew rostering plan will definitely be affected if we constantly assign crew routes with long working times to some crewmembers. Furthermore, our study of crew rostering problem on high-speed railway lines has the background that all crewmembers have different licenses. The income of

the crewmember per unit time is different because the license needed for finishing a crew route may be different, resulting in the divergence of the total income. In summary, the fairness and reasonableness of the plan can be improved by maintaining the balance of the working time and the income between all crewmembers. Thus, the objectives considered when establishing the optimization model of our problem are as follows.

- (a) To minimize the number of crewmembers used in the plan: in our research, all crew routes are divided into several rotas, which need the same number of crewmembers. Therefore, the number of rotas will be minimized if the number of crewmembers is minimized.
- (b) To minimize the differences in working time of all rotas: the working time balance of the crew rostering plan is equal to the working time balance among all rotas because the crewmembers' work in one rota is the same.
- (c) To minimize the differences in income of all rotas: the difference in license types is reflected by the income of the rotas because the assignment of crew routes to the fully trained crewmembers is not affected by the license type.

In working out the crew rostering plan, the following regulations and constraints should be satisfied.

- (a) A crew route should belong to one and only one rota.
- (b) For each rota, the number of tasks (crew route) contained in one day should not be more than one.
- (c) For each rota, the crew routes should not be more than the working days specified by the plan.
- (d) For each rota, the connection time between each two adjacent tasks must satisfy the connection standard set by the company.

According to the objectives and constraints analyzed above, the optimization model established in this paper is as follows:

$$\begin{aligned} \min \quad & Z_1 = M, \\ \min \quad & Z_2 = \sum_{i=1}^M \left( \sum_{k=1}^T \sum_{j=1}^N x_{ij}^k t_j - \frac{\sum_{j=1}^N t_j}{M} \right)^2, \end{aligned} \quad (1)$$

$$\begin{aligned} \min \quad & Z_3 = \sum_{i=1}^M \left( \sum_{k=1}^T \sum_{j=1}^N x_{ij}^k t_j l_j - \frac{\sum_{j=1}^N t_j l_j}{M} \right)^2, \\ \text{subject to} \quad & \sum_{i=1}^M \sum_{k=1}^{T_w} x_{ij}^k = 1 \quad j = 1, 2, \dots, N; \end{aligned} \quad (2)$$

$$\sum_{j=1}^N x_{ij}^k \leq 1 \quad i = 1, 2, \dots, M; \quad (3)$$

$$k = 1, 2, \dots, T_w;$$

$$\sum_{k=1}^{T_w} \sum_{j=1}^N x_{ij}^k \leq T_w \quad i = 1, 2, \dots, M; \quad (4)$$

$$\sum_{j=1}^N x_{ij}^{k+1} (t_j^s + 1440) - \sum_{j=1}^N x_{ij}^k t_j^e$$

$$+ \theta \left( 2 - \sum_{j=1}^N x_{ij}^{k+1} - \sum_{j=1}^N x_{ij}^k \right) \quad (5)$$

$$\geq T_{\text{con}} \quad i = 1, 2, \dots, M;$$

$$k = 1, 2, \dots, T_w - 1;$$

$$x_{ij}^k \in \{0, 1\}, \quad (6)$$

where  $M$  is the total number of rotas in the crew rostering plan;  $N$  is the total number of crew routes;  $T$  is the planning horizon of the crew rostering plan; and  $T_w$  is the working day of the plan;  $x_{ij}^k$  is a 0-1 decision variable equal to 1 if the  $k$ th crew route in rota  $i$  is crew route  $j$ , and 0 otherwise;  $t_j^s$ ,  $t_j^e$ , and  $t_j$  represent the starting, ending, and working times of crew route  $j$ , respectively;  $l_j$  is the income per unit time corresponding to the level of the license needed by crew route  $j$ ;  $\theta$  is a large positive number to ensure the establishment of formula (5); and  $T_{\text{con}}$  is the time standard that should be satisfied by each connection between two crew routes.

### 3.2. Model Analysis

**3.2.1. Analysis of the Objective Functions.** The established 0-1 integer programming model in our research consists of three objective functions.  $Z_1$  is the number of rotas that directly determines the number of crewmembers used in the crew rostering plan, while  $Z_2$  and  $Z_3$  are the objectives representing the fairness and reasonableness of the crew rostering plan, respectively. For a particular crew route, the income of a crewmember is proportional to its working time. For all crew routes, the crew route with long working time is not necessarily the one with high income, or vice versa, because of the difference in license type. Therefore, the optimal solutions of the objective functions  $Z_2$  and  $Z_3$  in the above model are hard to be achieved simultaneously. Objective function  $Z_2$  (working time equalization) can ensure that all the crewmembers have similar intensity of work at the planning horizon and good working status. Conversely, each crewmember hopes that her income is not below the average standard while her licenses and working time are the same as the others. Thus, the design of objective function  $Z_3$  helps maintain the stability of the crewmembers' emotion. However, the resting time and the working status of crewmembers are still favorable even without the consideration of  $Z_3$ . Therefore, the organization of the crew rostering plan is not likely to be

affected by  $Z_3$ . The above analysis indicates that the order of the importance of these objective functions in our model is shown as follows:

$$Z_1 \gg Z_2 > Z_3. \quad (7)$$

According to the approach for designing fitness function in Niu [25, 26], the model with multiobjectives in our research is transformed into a single objective optimization model according to the relationship between these objectives to evaluate the quality of the constructed solution. Objective function  $Z_1$  represents the number of rotas, which must be an integer. A new objective function can be designed instead of the original three objective functions by transforming  $Z_2$  and  $Z_3$  into a value between (0, 1):

$$Z = Z_1 + \lambda \frac{Z_2}{Z_2^*} + (1 - \lambda) \frac{Z_3}{Z_3^*}. \quad (8)$$

$Z_2^*$  and  $Z_3^*$  are the maximum values of the latter two objective functions theoretically.  $\lambda$  and  $(1 - \lambda)$  are their weights in the new objective function  $Z$ . The value of  $\lambda$  should be greater than 0.5 because  $Z_2$  is more important than  $Z_3$ :

$$Z_2^* = M[(\max\{t_j\} - \min\{t_j\})T_w]^2, \quad (9)$$

$$Z_3^* = M[(\max\{l_j t_j\} - \min\{l_j t_j\})T_w]^2,$$

where  $\max\{t_j\}$  and  $\min\{t_j\}$  are the longest and shortest working times of all crew routes, respectively, and  $\max\{l_j t_j\}$  and  $\min\{l_j t_j\}$  are the highest and lowest incomes of all crew routes.

**3.2.2. Analysis of the Solving Algorithm.** When only the working time of the crew routes is considered, the crew rostering problem described in our research is similar to a classical  $k$ -partitioning problem in complexity theory, regardless of the income and the connecting relationships between the crew routes. The  $k$ -partitioning problem is to find a partition  $S_1, S_2, \dots, S_m$  of a set  $\{I_1, I_2, \dots, I_n\}$  by satisfying the constraints  $w_j \geq 0$ ,  $|S_i| = k$ , and  $n = mk$  ( $w_j$  is the weight of  $I_j$  and  $|S_i|$  is the elements contained in subset  $S_i$ ) to minimize the maximum weight of each subset.

When the constraint of  $n = mk$  is relaxed, the elements in each subset will be  $\lfloor n/m \rfloor$  or  $\lceil n/m \rceil$ . This problem is called cardinality-constrained  $P| \cdot |C_{\text{max}}$  problem, which is NP-hard [27].

The  $k$ -partitioning problem is commonly used in scheduling theory to minimize the maximum working time when assigning  $n$  jobs that are independent of each other and no priority order to multiple processing systems [28]. An approximation algorithm used to solve this problem is the largest processing time (LPT) algorithm proposed by Graham. This algorithm initially sorts the tasks in descending order by their processing time and then assigns an unhandled task with the longest processing time to an idle processor. For a scheduling problem with  $m$  processors, the performance ratio of LPT compared with the optimal algorithm is  $(4/3) - (1/3m)$  [29]. The  $k$ -partitioning problem is a special

case of the cardinality-constrained  $P| \cdot |C_{\max}$  problem. The applicability of LPT to the cardinality-constrained  $P| \cdot |C_{\max}$  problem is proven by studies carried out by the follow-up scholars, and the worst performance ratio remains  $(4/3) - (1/3m)$  [30]. Although many scholars have studied the cardinality-constrained  $P| \cdot |C_{\max}$  problem, a unified and efficient algorithm for solving it has not yet been developed.

In this paper, the working time and income of crewmembers, as well as the connections between crew routes, are considered simultaneously in the optimization model, making this problem more complex than the cardinality-constrained  $P| \cdot |C_{\max}$  problem. Therefore, finding an optimal solution by directly solving the 0-1 integer programming model is difficult. The planner hopes that a feasible crew rostering plan can be obtained in a short time and that an optimal plan can be achieved by adjusting the feasible plan according to his/her working experience and other conditions. Therefore, a method that can rapidly generate a feasible plan is valuable.

## 4. Algorithms

The optimization objectives of our study are to minimize the number of crewmembers used and maintain the balance of working times and incomes among all crewmembers. A reasonable cycle scheme should be given before the establishment of crew rostering plan because the cycle schemes significantly affect the first objective. The method for deciding the cycle scheme and the method for working out the crew rostering plan under a given cycle scheme are represented as follows.

**4.1. Calculating the Optimal Cycle Scheme.** The monthly working time of the crewmembers is the foundation for deciding the cycle scheme, which means a reasonable proportion of working time and resting time in the cycle scheme is needed for crewmembers to finish the monthly working time. Meanwhile, the crewmembers were grouped in our research. Hence, the number of crewmembers is equal to the product of the number of rotas and the planning horizon of the plan. The crew routes in each rota should be as high as possible, which requires good proportions between the number of crew routes and the working days in the plan, to reduce the number of crewmembers used. Therefore, the optimal cycle scheme should be determined according to the number of crew routes, the average working time of crew routes, and the monthly working time standard.

By studying the relationship between the above factors, the method for deciding the optimal cycle scheme is designed as follows.

**4.1.1. Preliminary Selection for the Reasonable Cycle Scheme.** With the assumption that the number of crew routes is  $N$ , the average working time is  $t_{\text{average}}$  (min), the monthly working time standard is  $T_{\text{month}}$  (h), and the number of days in

a month is  $D_{\text{month}}$ . Then, the reasonable cycle scheme should satisfy

$$\frac{t_{\text{average}} \times T_{\text{work}}}{T} = \frac{60T_{\text{month}}}{D_{\text{month}}}, \quad (10)$$

where the values of  $T$  and  $T_{\text{work}}$  are unknown. However,  $T_{\text{work}}$  is a part of  $T$ ; thus, the relationship between them can be represented as  $\varphi$ , given by

$$\varphi = \frac{T_{\text{work}}}{T} = \frac{60T_{\text{month}}}{t_{\text{average}}D_{\text{month}}}. \quad (11)$$

Under the standard of China railway, the value of monthly working time is set to  $T_{\text{month}} = 166.7$  h and the day in a month to  $D_{\text{month}} = 30$  d. Then, formula (11) can be converted as follows:

$$\varphi = \frac{T_{\text{work}}}{T} = \frac{60 \times 166.7}{30t_{\text{average}}} = \frac{333.4}{t_{\text{average}}}. \quad (12)$$

Considering that  $T$  and  $T_{\text{work}}$  represent the number of days, the values of  $T$  and  $T_{\text{work}}$  are integers and their proportion close to  $\varphi$  is a reasonable cycle scheme. Alternatively, we can let  $T$  (or  $T_{\text{work}}$ ) be an integer and then calculate the value of  $T_{\text{work}}$  (or  $T$ ) by formula (12). When the calculated result is close to an integer, this cycle scheme is also a reasonable cycle scheme. A set of reasonable cycle schemes of the crew rostering plan can be drawn from this method, and then further selection will be executed among these cycle schemes.

**4.1.2. Further Selection for the Optimal Cycle Scheme.** With the assumption that the planning horizon of cycle scheme  $i$  is  $T^i$  and the number of working days is  $T_{\text{work}}^i$ , the number of crewmembers used in this crew rostering plan  $N_{\text{crew}}^i$  is calculated as follows:

$$N_{\text{crew}}^i = \left\lceil \frac{N}{T_{\text{work}}^i} \right\rceil \times T^i. \quad (13)$$

The cycle scheme with the minimal crewmembers is the optimal one in all reasonable cycle schemes because the number of crewmembers used in the plan is the most important objective. Formula (13) assumes that a crew route can connect to any other route in a short time, resulting in an ideal one. A *next day's connection* may exist between two crew routes because the connection time does not satisfy the regulations (which means adding 1440 min to their original connection time). This situation means that an empty crew route is added between two adjacent crew routes. When this situation occurs, the time in the plan cannot be used sufficiently and the optimal cycle scheme determined by formula (13) is improper. To fully consider the influence caused by the *next day's connections*, formula (13) is rewritten as

$$N_{\text{crew}}^i = \left\lceil \frac{N + \Delta n}{T_{\text{work}}^i} \right\rceil \times T^i. \quad (14)$$

TABLE 1: Crew route data.

Route index	Starting time (min)	Ending time (min)	Working time (min)	License type
1	364	806	442	1
2	370	737	367	3
3	371	825	454	3
4	394	840	446	2
5	405	801	396	3
6	430	819	389	2
7	442	815	373	2
8	457	846	389	3
9	512	954	442	1
10	527	920	393	1
11	542	942	400	2
12	580	1052	472	2
13	602	999	397	2
14	606	990	384	2
15	619	1036	417	3
16	624	1016	392	2
17	642	1093	451	2
18	645	1106	461	1
19	653	1037	384	2
20	678	1153	475	2
21	690	1090	400	2
22	703	1087	384	1
23	704	1155	451	2
24	707	1118	411	3
25	717	1192	475	1
26	718	1160	442	1
27	753	1190	437	2
28	756	1222	466	2
29	776	1240	464	1
30	784	1184	400	2
31	790	1252	462	3
32	808	1194	386	3
33	818	1246	428	2
34	824	1247	423	3
35	836	1301	465	2
36	836	1277	441	3
37	850	1253	403	1
38	854	1307	453	2
39	857	1305	448	2
40	883	1310	427	2
41	913	1373	460	2
42	916	1376	460	3
43	927	1372	445	2

*4.2. Improved Ant Colony Algorithm.* The method for solving the crew rostering problem in our research is then designed based on the adjustment of the ant colony algorithm with bi-pheromone and bi-heuristic information designed in the study by Tian and Niu [24]. The algorithm in this paper only differs in the design of heuristic information and evaluation function; thus, only these two parts will be described below.

For the other parts of the ant colony algorithm, interested readers can learn the related contents in the studies by Dorigo and Gambardella [31] and Tian and Niu [24].

*4.2.1. Calculation of Heuristic Information.* When constructing a rota, the calculation method of the selection probability

TABLE 2: Reasonable cycle schemes to be selected.

Cycle scheme	Working day	Planning horizon
1	1	1.28
2	2	2.56
3	3	<b>3.84</b>
4	4	<b>5.12</b>
5	5	6.40
6	6	7.68
7	7	<b>8.97</b>
8	8	10.25
9	9	11.53
10	10	<b>12.81</b>
11	11	<b>14.09</b>
12	12	15.37
13	13	16.65
14	14	<b>17.93</b>
15	15	19.21
16	16	20.49
17	17	21.77
18	18	<b>23.05</b>
19	19	24.33
20	20	25.61
21	21	<b>26.90</b>
22	22	<b>28.18</b>
23	23	29.46
24	24	30.74
25	25	<b>32.02</b>
26	26	33.30
27	27	34.58
28	28	<b>35.86</b>

is different according to the location of the ant. When the construction of a rota is finished, the ant returns to the virtual node, and then the first crew route for next rota is chosen by the pheromone and heuristic information of all crew route nodes. Otherwise, the next crew route is chosen by the pheromone and heuristic information of the connecting arcs if the constructing process is still searching for a crew route for current rota.

The heuristic information of crew route nodes is calculated by its starting time and ending time:

$$\eta_j = \frac{1}{(t_j^s + t_j^e)}. \quad (15)$$

The heuristic information of the connecting arcs between the crew routes should guide the ants finding a crew rostering plan that not only uses lower crewmembers but also keeps the balance of the working time and the income among all rotas. Therefore, the calculation of the heuristic information on arcs should contain the following parameters: (a) the connection time between crew routes, (b) the working time of the crew routes to be selected, and (c) the income of the crew routes to be selected.

When the connection time between two crew routes satisfies the regulations, the value of the time is as short as possible. Thus, the connection time is considered by its reciprocal in heuristic information. For the working time and income, we hope that the selection of the next crew route helps keep their balance in a final solution. Let  $T_{\text{now}}$  and  $I_{\text{now}}$  represent the working time and income of current crew route sequence, respectively, and  $d$  is the number of days in current sequence.  $T_{\text{average}}$  and  $I_{\text{average}}$  are the average working time and average income, respectively. The expected working time and income are  $dT_{\text{average}}/T_w$  and  $dI_{\text{average}}/T_w$ , respectively. The expected working time and income after the connection of next crew route are  $(d+1)T_{\text{average}}/T_w$  and  $(d+1)I_{\text{average}}/T_w$ , respectively. Therefore, the expected working time  $t_{\text{next}}$  and income  $r_{\text{next}}$  of next crew route to be selected are represented by the following formulas:

$$t_{\text{next}} = \frac{(d+1)T_{\text{average}}}{T_w} - T_{\text{now}}, \quad (16)$$

$$r_{\text{next}} = \frac{(d+1)I_{\text{average}}}{T_w} - I_{\text{now}}.$$

When constructing the rota, a crew route is more likely to be selected if its working time and income are close to  $t_{\text{next}}$  and  $r_{\text{next}}$ . Therefore, the heuristic information for selecting next crew route from current crew route is represented as

$$\eta_{ij} = \frac{1}{t_{ij}} \times \frac{1}{|t_i - t_{\text{next}}| + \sigma} \times \frac{1}{|r_i - r_{\text{next}}| + \sigma}, \quad (17)$$

where the last two parts denote the proportion of the working time and income of the crew routes to be selected and  $\sigma$  is a nonzero positive number that ensures that the value of  $\eta_{ij}$  is meaningful when  $t_i = t_{\text{next}}$  or  $r_i = r_{\text{next}}$ .

**4.2.2. Evaluation Function.** The integrated optimization objective (8) is used as the criteria for evaluating the quality of the solution constructed by the ants. According to the weight and range of all objectives, the value of function (8) consists of two parts, in which the integer part represents the number of rotas in the crew rostering plan, whereas the decimal part indicates the equilibrium of the working time and income of the rotas.

## 5. Numerical Example

To verify the reasonableness of the model and the effectiveness of the algorithm, this paper generates 43 crew routes shown in Table 1 to work out the crew rostering plan. The earliest starting time of crew routes is 6:04, whereas the latest finishing time is 22:56; when represented by minute, they are 364 and 1376, respectively. The working time of a crew route is the time span between its starting time and ending time. In our study, three license types have incomes of 0.55, 0.65, and 0.75 yuan per minute. The weights of the equilibrium of working time and income are set as 0.6 and 0.4.

TABLE 3: Optimal cycle schemes with different  $\Delta n$ .

Scheme	Reasonable cycle schemes		Crewmembers used under certain cycle scheme							
	Planning horizon	Working days	$\Delta n = 0$	$\Delta n = 1$	$\Delta n = 2$	$\Delta n = 3$	$\Delta n = 4$	$\Delta n = 5$	$\Delta n = 6$	$\Delta n = 7$
1	4	3	60	60	<b>60*</b>	64	64	64	68	68
2	9	7	63	63	63	<b>63*</b>	<b>63*</b>	<b>63*</b>	<b>63*</b>	72
3	13	10	65	65	65	65	65	65	65	65
4	14	11	<b>56*</b>	<b>56*</b>	70	70	70	70	70	70
5	18	14	72	72	72	72	72	72	72	72
6	23	18	69	69	69	69	69	69	69	69
7	27	21	81	81	81	81	81	81	81	81
8	28	22	<b>56*</b>	<b>56*</b>	84	84	84	84	84	84
9	32	25	64	64	64	64	64	64	64	<b>64*</b>
10	36	28	72	72	72	72	72	72	72	72

TABLE 4: Optimal crew rostering plan with  $T = 14$  and  $T_{work} = 11$ .

Rota	Crew route sequence											Working time (min)	Income (yuan)
1	2	31	6	22	7	37	12	16	13	5	32	4421	5912.1
2	8	9	17	43	20	30	40	11	14	3	18	4728	6134
3	10	23	38	26	39	19	21	1	33	25	15	4733	5886
4	4	24	28	42	35	41	27	29	34	36		4473	6069

Crewmembers used: 56, evaluation function: 4.0615, average working time: 4588.75 min, and average income: 6000.4 yuan.

According to the given data, the average working time of our 43 crew routes is 427 min. Thus, the proportion of  $T$  and  $T_w$  is

$$\varphi = \frac{T_{work}}{T} = \frac{333.4}{427} = 0.78. \tag{18}$$

To use the crewmembers more flexibly, the planning horizon of long period is not considered in our research. To determine the optimal value of  $T$  and  $T_{work}$ , we let  $T_{work}$  be an integer and then calculate the value of  $T$ . When  $T_{work}$  belongs to  $[1, 28]$ , the values of  $T$  calculated by  $\varphi$  are shown in Table 2, in which the bold figures mean that the values of  $T$  are close to integer.

Furthermore, when taking the number of crew routes  $N$  and the number of probable *next day's connection*  $\Delta n$  into account, the best cycle schemes are shown in Table 3.

We can conclude from Table 3 that the optimal cycle scheme varies with the number of the *next day's connection* in the crew rostering plan, which causes the number of the crewmembers used in the plan to vary. The number of *next day's connection* is unpredictable; thus, we choose the index of the cycle scheme by the order of 4(8)-1-2-9 while  $\Delta n$  varies from 0 to 7, as shown in Table 3 with asterisks. If a crew rostering plan using 56 crewmembers can be obtained under the 4th (or 8th) value of  $T$  and  $T_w$ , we believe that the plan is already the best one. Otherwise, the next cycle scheme will be selected to work out the crew rostering plan again.

After adopting the fourth and eighth cycle schemes in Table 3, the optimal crew rostering plan shown in Tables 4 and 5 is obtained by the improved ant colony algorithm designed in 4.2. The parameters of the improved ant colony algorithm are set to  $\alpha = 1$ ,  $\beta = 5$ , and  $\rho = 0.1$ ; the number of ants is 30; and the iteration time is 100. The algorithm

is programmed by Visual C# on a personal computer with 2.93G CPU and 1 G memory. The average operation times for the two different cycle schemes are 1.34 and 1.27 s.

Tables 4 and 5 show that the number of rotas is 4 when  $T = 14$  and  $T_w = 11$  and 2 when  $T = 28$  and  $T_w = 22$ . The numbers of crewmembers used in two plans are all 56, which reaches the lowest bound. The result shows that the two cycle schemes are optimal and helpful to reduce the number of crewmembers. The values of evaluation function of the two cycle schemes are shown in the tables.

## 6. Conclusions

With the development of China high-speed railway lines, new changes have occurred in the organization of operation plans. As an important part of CPP, the research of crew rostering problem is crucial for optimizing the use of crewmembers. In this paper, the multilicense crew rostering problem is studied by adopting the grouped rostering approach, which is helpful to the operation and management of the crewmembers. The optimization model aimed at minimizing the number of crewmembers used and keeping the balance of the working time and income of all crewmembers has been established. In our research, we found that the cycle scheme is the main factor that influences the number of crewmembers used in the plan. Then, the calculation method of the optimal cycle scheme is studied with the consideration of the number of crew routes, the average working time of crew routes, and the monthly working time standard of crewmembers. After the cycle scheme is selected, an ant colony algorithm with bi-pheromone and bi-heuristic information is designed to work out the crew rostering plan. In the solving process,

TABLE 5: Optimal crew rostering plan with  $T = 28$  and  $T_{\text{work}} = 22$ .

Rota	Crew route sequence											Working time (min)	Income (yuan)
1	2	10	8	21	1	4	17	23	3	7	22	9106	11838.8
	37	13	6	38	16	19	40	15	35	14	43		
2	5	9	25	32	33	18	24	41	31	20	26	9249	12162.7
	28	12	11	36	30	39	34	27	29	42			

Crewmembers used: 56, evaluation function: 2.0035, average working time: 9177.5 min, and average income: 12000.8 yuan.

the algorithm can dynamically calculate the expected working time and income for choosing the next crew route according to the accumulative working time and income of the partial solution of the constructed rota. In this way, the algorithm can select a reasonable crew route that meets the objective function of the optimization model. The designed numerical example showed that the designed algorithm can select an optimal cycle scheme, and the crew rostering plan weaved with this cycle scheme also met the optimization goals of the problem.

### Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

### Acknowledgment

This research work was supported by the National Natural Science Foundation of China (no. 71261014).

### References

- [1] N. Kohl and S. E. Karisch, "Airline crew rostering: problem types, modeling, and optimization," *Annals of Operations Research*, vol. 127, no. 1–4, pp. 223–257, 2004.
- [2] A. M. Cohn and C. Barnhart, "Improving crew scheduling by incorporating key maintenance routing decisions," *Operations Research*, vol. 51, no. 3, pp. 387–396, 2003.
- [3] S. Jütte and U. W. Thonemann, "Divide-and-price: a decomposition algorithm for solving large railway crew scheduling problems," *European Journal of Operational Research*, vol. 219, no. 2, pp. 214–223, 2012.
- [4] C. Barnhart, L. Hatay, and E. L. Johnson, "Deadhead selection for the long-haul crew pairing problem," *Operations Research*, vol. 43, no. 3, pp. 491–495, 1995.
- [5] G. W. Graves, R. D. McBride, I. Gershkoff, D. Anderson, and D. Mahidhara, "Flight crew scheduling," *Management Science*, vol. 39, no. 6, pp. 736–745, 1993.
- [6] A. T. Ernst, H. Jiang, M. Krishnamoorthy, and D. Sier, "Staff scheduling and rostering: a review of applications, methods and models," *European Journal of Operational Research*, vol. 153, no. 1, pp. 3–27, 2004.
- [7] T. Emden-Weinert, H. Kotas, and U. Speer, "DISSY-A Driver Scheduling System for Public Transport. Version 1.0," 2001, <http://wenku.baidu.com/view/4e960e0d76c66137-ee06196d.html>.
- [8] M. Gamache, F. Soumis, G. Marquis, and J. Desrosiers, "Column generation approach for large-scale aircrew rostering problems," *Operations Research*, vol. 47, no. 2, pp. 247–262, 1999.
- [9] S. C. K. Chu, "Generating, scheduling and rostering of shift crew-duties: applications at the Hong Kong International Airport," *European Journal of Operational Research*, vol. 177, no. 3, pp. 1764–1778, 2007.
- [10] J. Desrosiers, A. Lasry, D. McInnis, M. M. Solomon, and F. Soumis, "Air transat uses ALTITUDE to manage its aircraft routing, crew pairing, and work assignment," *Interfaces*, vol. 30, no. 2, pp. 41–53, 2000.
- [11] B. Maenhout and M. Vanhoucke, "A hybrid scatter search heuristic for personalized crew rostering in the airline industry," *European Journal of Operational Research*, vol. 206, no. 1, pp. 155–167, 2010.
- [12] F.-Y. Chu, Z.-Q. Tian, and S.-Q. Ni, "Model and algorithm for formulation of the single cycle crew rostering plans of high-speed railway," *Journal of the China Railway Society*, vol. 34, no. 7, pp. 1–9, 2012.
- [13] A. Corominas, R. Pastor, and E. Rodríguez, "Rotational allocation of tasks to multifunctional workers in a service industry," *International Journal of Production Economics*, vol. 103, no. 1, pp. 3–9, 2006.
- [14] A. N. Avramidis, W. Chan, and P. L'Ecuyer, "Staffing multi-skill call centers via search methods and a performance approximation," *IIE Transactions*, vol. 41, no. 6, pp. 483–497, 2009.
- [15] J. Buhr, "Four methods for monthly crew assignment—a comparison of efficiency," in *Proceedings of the AGIFORS Symposium*, vol. 18, pp. 403–430, 1978.
- [16] N. Beaumont, "Using mixed integer programming to design employee rosters," *The Journal of the Operational Research Society*, vol. 48, no. 6, pp. 585–590, 1997.
- [17] D. M. Ryan, "The solution of massive generalized set partitioning problems in air crew rostering," *The Journal of the Operational Research Society*, vol. 43, pp. 459–467, 1992.
- [18] D. Teodorović and P. Lučić, "A fuzzy set theory approach to the aircrew rostering problem," *Fuzzy Sets and Systems*, vol. 95, no. 3, pp. 261–271, 1998.
- [19] P. Lučić and D. Teodorović, "Simulated annealing for the multi-objective aircrew rostering problem," *Transportation Research A*, vol. 33, pp. 19–45, 1999.
- [20] H. Dawid, J. König, and C. Strauss, "An enhanced rostering model for airline crews," *Computers and Operations Research*, vol. 28, no. 7, pp. 671–688, 2001.
- [21] A. Ernst, M. Krishnamoorthy, and D. Dowling, "Train crew rostering using simulated annealing," in *Proceedings of the International Conference on Optimization: Techniques and Applications (ICOTA '98)*, Perth, Australia, 1998.
- [22] B. Santosa, A. Sunarto, and A. Rahman, "Using differential evolution method to solve crew rostering problem," *Applied Mathematics*, vol. 1, no. 4, pp. 316–325, 2010.
- [23] A. Monfroglio, "Hybrid genetic algorithms for a rostering problem," *Software Practices Expert*, vol. 26, no. 7, pp. 851–862, 1996.

- [24] Z. Tian and H. Niu, "Modeling and algorithms of the crew rostering problem with given cycle on high-speed railway lines," *Mathematical Problems in Engineering*, vol. 2012, Article ID 214607, 15 pages, 2012.
- [25] H. Niu and M. Zhang, "An optimization to schedule train operations with phase-regular framework for intercity rail lines," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 549374, 13 pages, 2012.
- [26] H. Niu and X. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transportation Research C*, vol. 36, pp. 212–230, 2013.
- [27] M. Dell'Amico and S. Martello, "Bounds for the cardinality constrained  $P||C_{max}$  problem," *Journal of Scheduling*, vol. 4, no. 3, pp. 123–138, 2001.
- [28] K. R. Baker, *Introduction to Sequencing and Scheduling*, John Wiley & Sons, New York, NY, USA, 1974.
- [29] R. L. Graham, "Bounds on multiprocessing timing anomalies," *SIAM Journal on Applied Mathematics*, vol. 17, no. 2, pp. 416–429, 1969.
- [30] H. Kellerer and G. Woeginger, "A tight bound for 3- partitioning," *Discrete Applied Mathematics*, vol. 45, no. 3, pp. 249–259, 1993.
- [31] M. Dorigo and L. M. Gambardella, "Ant colonies for the travelling salesman problem," *BioSystems*, vol. 43, no. 2, pp. 73–81, 1997.

## Research Article

# Modeling the Perceptions and Preferences of Pedestrians on Crossing Facilities

Hongwei Guo,<sup>1</sup> Facheng Zhao,<sup>1</sup> Wuhong Wang,<sup>1</sup> Yanlong Zhou,<sup>1</sup>  
Yujie Zhang,<sup>1</sup> and Geert Wets<sup>2</sup>

<sup>1</sup> Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

<sup>2</sup> Transportation Research Institute (IMOB), Hasselt University, Wetenschapspark 5 bus 6, 3590 Diepenbeek, Belgium

Correspondence should be addressed to Wuhong Wang; wangwuhong@bit.edu.cn

Received 17 November 2013; Accepted 16 January 2014; Published 6 March 2014

Academic Editor: Huimin Niu

Copyright © 2014 Hongwei Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pedestrian's street-crossing behaviour has a significant effect on traffic performance and safety. The crossing behaviour is determined by human factors and environmental factors. Aiming at examining the pedestrian perceptions toward crossing facilities and preferences for crossing locations, an observational study of pedestrian crossing behaviour at urban street is conducted. The perceptions and preferences of pedestrians are collected using stated preference technique. A specific questionnaire is designed to conduct the stated preference survey. A multinomial logit model is proposed to describe the perceptions and preferences of pedestrians on crossing facilities and locations. The sensitivity analysis is performed to discuss the influence of various factors on crossing behaviour. Then the relationship between crossing locations and crossing distances is analyzed by a new proposed method. With the theoretical analysis, the engineering solutions considering pedestrian behaviour are suggested. The results are helpful to design human-centered crossing facilities in urban traffic.

## 1. Introduction

Pedestrians' traffic behaviour at urban network includes walk along streets and cross streets. Once pedestrians want to cross streets, it is inevitable that they will conflict with motor vehicles. Traffic accidents that involved pedestrians and cyclists have become a critical safety problem all over the world [1]. In a developing country, like China, with large population and weak infrastructure, pedestrians often become a reason for traffic congestion and traffic accidents. To deal with pedestrian traffic problems, various crossing facilities are designed to assist pedestrian in crossing safely, for example, crosswalk (signalized and unsignalized), pedestrian overpass, and pedestrian underpass at intersection or midblock. With crossing facilities, pedestrians are separated from motor vehicles temporally or spatially. Unfortunately, pedestrians' crossing behaviour is strongly related to human factors. Thus pedestrians may cross illegally rather than using crossing facilities. The subjectivity and randomness make pedestrian behaviour complicated and also encourage traffic engineers to pay more attention to pedestrian traffic.

The research on pedestrian characteristics is the basic and important part for traffic engineering. Fruin [2] published the monograph *Pedestrian Planning and Design* and it is regarded as a foundation for pedestrian research. Then some particular researches included studies for behaviour, psychology, safety, and simulation. Mohammed [3] evaluated pedestrian crossing speed in Jordan and evaluated the effect of age, gender, and distance crossed (street width). Lam et al. [4] studied the relationship between walking speed and pedestrian flow under various flow conditions and the effects of bidirectional pedestrian flow on signalized crosswalks in Hong Kong. Sisiopiku and Akin [5] presented findings from an observational study of pedestrian behaviour at various urban crosswalks in a divided urban street near university campus. Pedestrian level-of-service assessment models were introduced by different methods under various traffic conditions, and standards for each level were estimated [6, 7].

Pedestrian behaviour is very complex and easily influenced by environmental designs and urban forms. A proper design of facilities can encourage walking without compromising safety and convenience [8, 9]. The waiting time

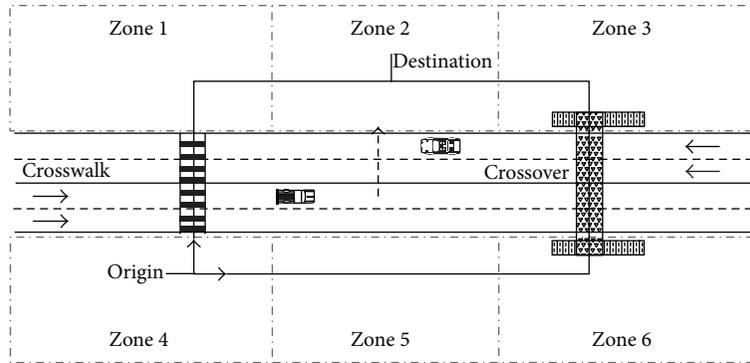


FIGURE 1: Illustration of crossing route.

and crossing distance (distance between trip destination and actual crossing location) are the mainly external factors which would lead to unsafe crossing. The need to hurry or the desire to keep moving along the shortcut is the main subjective reason behind the lack of compliance with pedestrian signals or crossing facilities. Pedestrian violation can be considered as the inevitable outcome of the contradiction between external factors and human factors. Li [10] and Lambrianidou et al. [11] gave various reports about pedestrians' behaviour influenced by time and distance. Guo et al. [12, 13] analysed the waiting behaviour during the street crossing by using the reliability theory and indicated the violation risk quantitatively with the increasing waiting time. Chu et al. [14] used data obtained from pedestrians' stated crossing preference and explained the stated preference with the street environment within the framework of disaggregate models. Yannis et al. [15] improved Chu's model to evaluate accident risk along a trip in relation to the estimated crossing behaviour of pedestrians. Nassiri and Sajed [16] evaluated and identified the effective parameters in pedestrian's decision-making process based upon vehicle speed and headway on multilane streets by using logit model.

It comes to a conclusion that a series of studies have provided insight into several aspects of pedestrian crossing behaviour. Compared with the researches about pedestrian accident analysis and safety evaluation, the researches about the preferences and perceptions of pedestrians on crossing facilities and crossing locations are limited. While crossing a street, the crossing facilities and crossing locations play an important role. The existence of crossing facilities is to ensure safety and assist accessibility. However, some pedestrians do not like using crossing facilities and even cross street illegally because they do not think the facilities can meet their demands. This paper aims at evaluating pedestrian preferences for commonly encountered crossing facilities and analysing pedestrian crossing locations in different conditions. To achieve such objective, a multinomial logit model for pedestrian crossing behaviour is proposed to analyse the probability distribution of utilities for various crossing facilities. With the evaluated model, pedestrian crossing locations are examined in various assumed traffic conditions so as to explore the reasons behind the crossing behaviour. This paper

hopes to give a better understanding of pedestrian crossing behaviour.

The paper is organized as follows. Section 2 introduces the methodology including discrete choice model and data collection. Section 3 describes the model demonstration and estimation. Section 4 discussed pedestrian crossing behaviour under various influential factors. Section 5 presents some engineering improvement for pedestrian facilities considering the crossing behaviour. Section 6 presents some conclusions from this study and a tentative plan for future work.

## 2. Methodology

*2.1. Analysis of Pedestrian Crossing Behaviour.* As mentioned above, pedestrian crossing behaviour is strongly related to human factors and traffic circumstances. Pedestrians' decision making about where to cross or when to cross can be described as the process of perception-judgment-decision-action. Generally, crossing decision is influenced by characteristics of the trip (e.g., the origin and destination and the complexity and the length of the route), characteristics of the infrastructure (e.g., the type of pedestrian facility, road geometry, and traffic conditions), and individual characteristics (e.g., age and gender and safety awareness). The crossing behaviour shall therefore reflect the combined assessment of the above characteristics under all sorts of conditions. In accordance with human nature, crossing behaviour exhibits significant subjectivity and randomness. Therefore, pedestrian crossing behaviour may become risk-taking action and lead to conflicts with motor vehicles.

The process of street crossing can be explained by the utility maximization theory that pedestrians want to choose the most satisfactory facilities and locations to cross the street. As a result, pedestrians get the maximum utility. It would be reasonable to assume that pedestrians' most satisfactory decisions are dependent on the location and the type of crossing facility. For example, as shown in Figure 1, there are two crossing facilities in the area. If a pedestrian's origin is in zone 5 and the destination is in zone 2, there are three potential routes to cross the street. The first route is to use

the crosswalk, the second route is to use the overpass, and the third route is to cross at any location. Therefore the existence of crossing facility may change pedestrians' crossing behaviour and also induce traffic violation because pedestrians cannot cross the street at the desired location.

**2.2. Discrete Choice Model Framework.** The utilitarian approach of microeconomic concepts is employed and the discrete choice model framework is adopted to describe crossing behaviour. The discrete choice model is one of the most important models for the research on traffic behaviour and is widely used in transportation predictions [17]. According to the stochastic utility model, various alternatives have utilities which will influence the choice. In the behaviour of facility choice, the alternative  $i$  (i.e., the crossing facility  $i$ ) within set of all elemental alternatives  $A_I$ ; the utility function  $U_{in}$  of facility  $i$  selected by pedestrian  $n$  consists of an observable utility component  $V_{in}$  and the unobservable random component  $\varepsilon_{in}$ . The utility can be defined as

$$U_{in} = V_{in} + \varepsilon_{in} \quad \forall i \in A_I; i = 1, \dots, I; n = 1, \dots, N. \quad (1)$$

Assuming that each pedestrian is the decision maker of him/herself, the selection criterion is to maximize the utility. In the stochastic utility model, the probability to choose crossing facility  $n$  is

$$\begin{aligned} P_{in} &= \text{Prob}(U_{in} \geq \max_{k \in A_I, k \neq i} U_{kn}) \\ &= \text{Prob}(V_{in} + \varepsilon_{in} \geq \max_{k \in A_I, k \neq i} (V_{kn} + \varepsilon_{kn})), \end{aligned} \quad (2)$$

where  $P_{in}$  is the probability of crossing facility  $i$  selected by pedestrian  $n$ ,  $U_{kn}$  is the utility function of other alternatives in choice set  $A_I$  excluding the alternative  $i$ , and  $\text{Prob}(\ast)$  is the probability function.

Operational models are based on specific assumptions about the distribution of  $\varepsilon_{in}$ . Assuming i.i.d. extreme value distributions lead to the multinomial logit (MNL) model, which has been very successful due to its computational and analytical tractability. The MNL model for the selection behaviour of pedestrian is defined as

$$P_{in} = \frac{\exp(V_{in})}{\sum_{j \in A_I} \exp(V_{jn})}. \quad (3)$$

**2.3. Data Collection.** Data collection was taken in specific area with more than two types of crossing facilities. The spacing between two facilities was within 300 meters. Moreover, there were large pedestrians to cross the street. The survey area was divided into 6 zones (or 8 zones) so that the pedestrians' origins and destinations can be recorded accurately. Finally, the Zhongguancun Street and Xidan Street in Beijing were chosen as survey areas.

The data collection contained field survey and questionnaire survey. The field survey obtained the information about pedestrians crossing behaviour. The field survey was conducted by observers at the appointed site and it contained two aspects: (a) to observe the crossing locations of pedestrians who had participated in the questionnaire survey and (b) to

observe and record the amounts of pedestrians used crossing facilities. Questionnaire survey was conducted to collect stated preference (SP) data. The design of the questionnaire should meet preset criteria such as inclusion of the statement of the study purpose and importance, clear definition of questions, and avoidance of personal or potentially offensive questions. The contents of questionnaire contained (a) pedestrians' personal profile (age and gender), (b) preference for crossing facilities (crossing location, detour distance, and compliance conditions), (c) action principle of crossing choice (i.e., which one is considered to be the first importance: safety, convenience, saving time, or saving strength?), (d) jaywalking conditions (i.e., reason and frequency), and (e) attitude towards detour (i.e., acceptance level and acceptable distance). The questionnaire was pretested before real survey. The whole survey for each pedestrian was finished in real traffic condition and it took less than 2-3 minutes. The information about the data collection and corresponding definition of variables in the survey are shown in Table 1.

At the end of the survey, 402 available questionnaires were received. There were 205 male pedestrians and 197 female pedestrians participating in the survey. The most participants were under 40 years of age. In the field survey, 1158 pedestrians' crossing locations and routes were recorded. These records also included the pedestrians who participated in the questionnaire survey. From the survey results, a majority of pedestrians used overpass/underpass to cross (46.4% used overpass, 32.7% used crosswalk, and 20.9% jaywalked) and most of them stated that they preferred overpass/underpass in questionnaire (42.1% choose overpass and 22.3% choose underpass). The remaining 35.6% preferred to cross at pedestrian crosswalk. About the action principle of crossing choice, 50.49% of participants thought of safety as the principle of crossing, 32.17% chose convenience, and 17.32% chose saving time or physical strength. The attitude towards the detour indicated that 30.6% of participants were willing to detour to use the crossing facility and 6.7% of them reject to detour. The remaining pedestrians stated they would accept detour in certain conditions (i.e., 37.3% of them detour sometime and 25.5% of them detour occasionally). Other results showed that most pedestrians (86.13%) could accept a detour distance less than 100 meters. And the results indicated that people are willing to detour in subjective desire but the detour distance they can accept is short relatively.

The survey results show the pedestrians' preference for crossing facilities. And the SP data are basically accordant with the RP data so that a conclusion for the survey can be drawn: the survey is available.

### 3. Model Demonstration and Estimation

**3.1. Model Demonstration.** According to the stochastic utility theory, the alternatives have various utilities that will influence the choice. In this paper, three alternatives are considered, including pedestrian crosswalk, overpass (underpass), and jaywalking. The overpass and underpass have the similar physical structure and safeguard so they are classified as the same alternative. Jaywalking means a pedestrian does

TABLE 1: Definition of variables.

Variable	Definition	Note
$X_{n1}$	Age	Age group: (1) 15–23, (2) 24–30, (3) 31–40, (4) 41–55, and (5) >55
$X_{n2}$	Gender	Pedestrian's gender: (1)—male and (2)—female
$X_{n3}$	Principles of crossing	The principles of street crossing: (1) safety, (2) convenience, and (3) save time or strength
$X_{n4}$	Detour willingness <sup>a</sup>	Pedestrian's attitude towards detour: (1) accept to detour, (2) often detour, (3) occasionally detour, and (4) refuse to detour
$X_{n5}$	Compliance with the traffic rules	The degree of compliance with the traffic rules: (1) always, (2) often, and (3) seldom
$X_{n6}$	Crossing time	The time for crossing at crosswalk: (1) pedestrian green signal, (2) pedestrian red signal and force to cross, and (3) pedestrian red signal and the road is clear
$X_{n7}$	Illegal reason	The reason behind the illegal crossing: (1) unreasonable design and (2) pedestrians' subjective reasons
$X_{n8}$	Conformity psychology	(1) Pedestrians follow others to cross street (2) Pedestrians do not follow others
$X_{n9}$	Detour distance	The additional distance caused by detour
$X_{n10}$	Travel time	The time spent on crossing street via crossing facility
$X_{n11}$	Origin and destination	The origin and destination of the trip of a pedestrian

<sup>a</sup>Detour means a pedestrian has to walk added distance to cross the street at designated location (facility) rather than to cross directly or at the desired location. The routes 1 and 2 are detour routes as shown in Figure 1.

TABLE 2: Model demonstration.

Variable	Alternative		
	Overpass/underpass	Crosswalk	Jaywalking
constant		●	●
$X_{n1}$	●		
$X_{n2}$		●	
$X_{n3}$	●		
$X_{n4}$	●		
$X_{n5}$			●
$X_{n6}$		●	
$X_{n7}$			●
$X_{n8}$		●	
$X_{n9}$			●
$X_{n10}$	●	●	●
$X_{n11}$	●	●	●

not cross the street using crossing facilities. Additionally, the model is based on the hypotheses that the pedestrian himself is the decision maker of crossing the street and he always chooses the satisfactory facilities. With the alternatives and explanatory variables adopted above, utility functions for each alternative can be defined. The variables and components of utility functions are shown in Table 2.

**3.2. Model Evaluation.** With the model evaluation, the parameters for the explanatory variables are estimated in each utility function by fitting the models to the observed choice data. Maximum likelihood estimation and  $t$ -test are adopted.

When the absolute value of  $t$ -test of one explanatory variable is greater than 1.96, this variable will be considered that it can influence the choice with the probability of 95%. If there is any variable that cannot influence the result significantly, it will be excluded and the other parameters will be estimated again. From initial estimation, the pedestrian's gender (variable  $X_{n2}$ ) and pedestrian's compliance with the traffic rules (variable  $X_{n6}$ ) are rejected. Then the remainder variables are estimated again and the results are shown in Table 3. With the results, a conclusion can be drawn that the model is well behaved. First, all explanatory variables can influence the choice significantly. Second, the variables fit the data well that the  $\rho^2$  adjusted for the number of variables is 0.396. In contrast, it is common to see an adjusted  $\rho^2$  below 0.3 in discrete choice model such as mode choice models [14].

**3.3. Validating Model Credibility.** The estimated model is applied with survey data to compare with the actual observations to establish the credibility of choice models and to evaluate whether or not the model demonstrates real behaviour. The alternative with maximum probability will be chosen. If a calculation result is accordant with the survey data, this is defined as a hit result. A hit result means a pedestrian's behaviour is modeled exactly. Let  $S_{in}$  represent the hit result that pedestrian  $n$  selects facility  $i$ . Consider the following equation:

$$S_{in} = \begin{cases} 1; & \text{if the calculated result is equal to} \\ & \text{the survey result} \\ 0; & \text{otherwise.} \end{cases} \quad (4)$$

TABLE 3: Coefficient estimation for explanatory variable.

Coefficient	$X_{n1}$	$X_{n3}$	$X_{n4}$	$X_{n5}$	$X_{n6}$	$X_{n7}$	$X_{n8}$	$X_{n9}$
Estimation	0.9112	-0.5828	-0.906	-0.4792	-2.341	-0.7021	-0.014	-0.0523
<i>t</i> -test	-2.3632	7.8355	-4.3234	-2.3632	-4.6136	4.2314	-2.8926	-2.2793

TABLE 4: Hit ratio of MNL model.

Alternative	Observed result	Calculation result	Accordant result	Hit ratio (%)
Crosswalk	131	131	97	74.05%
Overpass	187	187	154	82.44%
Jaywalk	84	84	59	70.23%
Total	402	402	310	77.11%

The hit ratio can be calculated as

$$R_i = \frac{\sum_{i=1}^3 S_{im}}{N_i}, \tag{5}$$

$$R = \frac{\sum_{i=1}^3 \sum_{m=1}^3 S_{im}}{\sum_{m=1}^3 J_m},$$

where  $R_i$  is hit ratio of alternative  $i$ ,  $R$  is hit ratio of all alternatives,  $N_i$  is the number of samples belonging to alternative  $i$ , and  $J_m$  is the number of entire samples.

The results of the hit ratio can be seen in Table 4.

From the results in Table 4, the hit ratio has indicated that the model can explain the preference of crossing facilities preferably. With the proposed model, the crossing behaviour can be discussed in detail.

#### 4. Discussion of Pedestrian Crossing Behaviour

With the aim of analyzing the crossing behaviour at various pedestrian facilities, pedestrians' crossing behaviour is discussed under specific hypothetical conditions. Here, sensitive analysis is performed using the proposed model to illustrate the impact of the considered variables, for example, action principles, detour willingness, and distance. These analyses are important to understand pedestrian behaviour.

*4.1. Effect of Action Principle.* When pedestrians want to cross the street, the action principle will influence the crossing choice significantly. If pedestrians consider safety the first importance while crossing street, they probably choose crossing facilities instead of illegal crossing. On the other hand, if the convenience is considered the first importance, jaywalking is easy to happen. This phenomenon is verified by the survey data: 41% of the samples consider safety the first importance and 39% have chosen the overpass/underpass. In order to study the influence of the action principle, three hypotheses are used: (a) all pedestrians choose safety as action principle, (b) convenience is chosen, and (c) saving time or physical strength is chosen. Pedestrians' crossing choices are evaluated by MNL model and the results are shown in Figure 2. In the results of hypothesis (a),

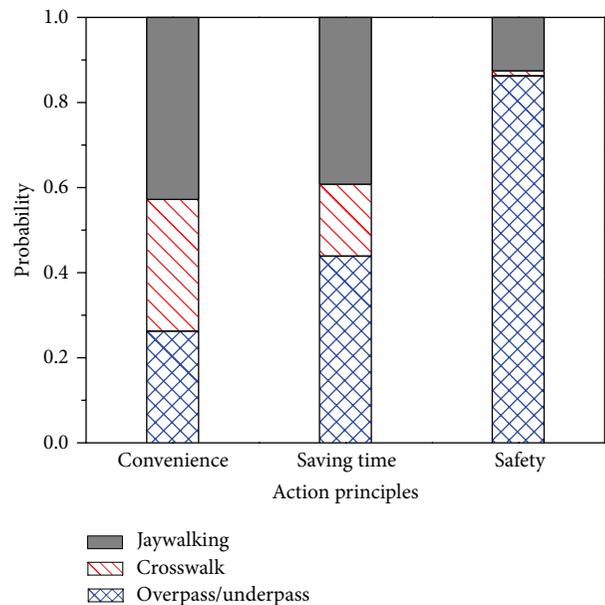


FIGURE 2: Probability of each alternative with various action principles.

the overwhelming majority of pedestrians (86.0%) will choose overpass/underpass to cross the street and only 1.2% of pedestrians will choose crosswalk. The results agree with the truth that overpass/overpass can provide better safeguard for pedestrians. Additionally, there are still a certain number of jaywalkers (12.8%) but this is not contrary to the model: not all pedestrians want to use crossing facilities. In the results of hypothesis (b), convenience is the action principle and the results show that almost half of the pedestrians (43.6%) will cross illegally. This is accordant with the current situation that jaywalking may be more convenient than crossing at specified facilities. Sometimes, using pedestrian facilities means increasing walking distance or waiting time. Pedestrians who choose the crosswalks (31.6%) are more than people who choose the overpass/underpasses (26.8%). The results of hypothesis (b) are similar to those of hypothesis (c). Jaywalking may be a proper approach to saving time or strength to a certain degree, although it is illegal.

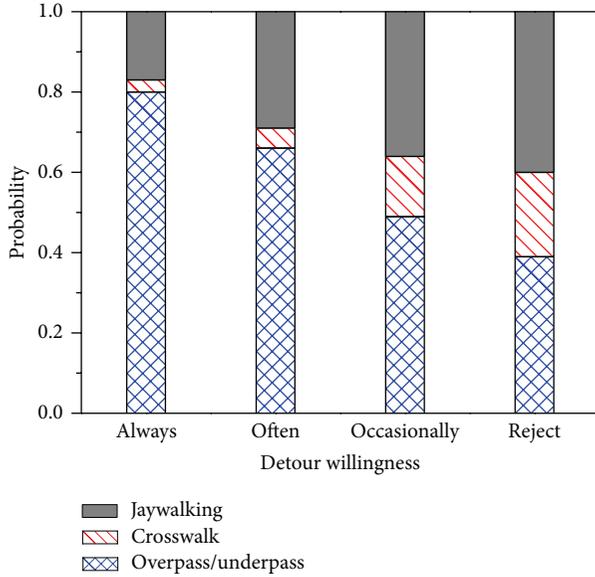


FIGURE 3: Probability of each alternative with different detour willingness.

**4.2. Effect of Detour Willingness and Detour Distance.** In real traffic condition, pedestrians may detour to arrive at the location with crossing facilities rather than crossing anywhere or anytime they want. Such detour increase crossing distance and influence crossing behaviour significantly. Therefore, it is important for engineers to design crossing facilities with reasonable interval distance. Here, pedestrians' attitudes about the detour and acceptable detour distance are considered. According to the questionnaire survey, pedestrians' attitudes towards the detour are evaluated by the frequency of detour, which is divided into four levels: always accept, often, occasionally, and reject. Here the crossing behaviour is represented under assumed conditions if all of the pedestrians select the same attitude to cross the street. From the results, as shown in Figure 3, with decrease of the acceptance frequency, the probability that the overpass/underpass is chosen is in a downtrend. In contrast, the probabilities that the crosswalk or jaywalking is chosen increases and the probability of jaywalking would increase even faster. These results are accordant with the reality that jaywalking is a direct and convenient way if pedestrians are not willing to detour to use the crossing facilities. Furthermore, the survey data also confirm the result (only 30.6% of the participants can accept detour and nearly half of them seldom detour).

On the other hand, the acceptable distance for detouring is important to explain the crossing behaviour that people cannot cross at the expected location. Pedestrians' attitudes towards detour would reflect their personal desire, but the detour distance may change the final choice. The survey data present the fact that majority of participants (86.13%) are willing to accept the detour distance within 100 meters. It means, if the detour distance is longer than 100 meters, pedestrians may change their mind even jaywalk. With the help of MNL model, the sensitivity about detour distance is analyzed. As shown in Figure 4, it is clear that the detour distance can influence the probability of jaywalking.

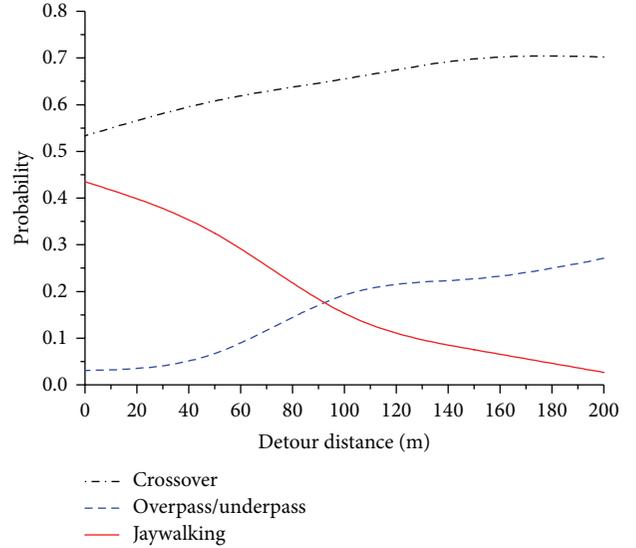


FIGURE 4: Probability of each alternative with different acceptable detour distance.

The results show that only 2.74% of people will cross illegally on the premise that acceptable distance is 200 meters. On the contrary, 43.5% of pedestrians may jaywalk if such distance is shorter than 50 meters.

Here, a parameter named compliance index (CI) is defined as the difference that the probability of legal crossing minus jaywalking probability. This parameter indicates the level to which pedestrians comply with traffic regulations when they are in conditions of various detour distances. So larger CI parameter means only a few pedestrians violate the traffic regulations. As shown in Figure 5, the CI curve decreases with the increase of detour distance. A rapid decrease appears between 75 and 150 meters. This trend is accordant with the survey results and it is useful to give some suggestions for design of pedestrian facilities. According to the curve fitting, the function of CI is given by

$$D_i(S) = \alpha \cdot \tanh\left(\frac{S}{\varphi} - \gamma\right) + \beta, \quad (6)$$

where  $D_i(S)$  is function of CI at location  $i$  and  $S$  is detour distance to get location  $i$ .

The CI curve is shown as a blue line and the fitting curve of (6) (with  $\alpha = 0.38$ ,  $\beta = 0.45$ ,  $\varphi = 46.66$ , and  $\gamma = 2.45$ ) is shown as a red line in Figure 5. As shown in Figure 5, (6) has a fair goodness of fit and it can present the features of CI curve.

**4.3. Walk First or Cross First.** The above discussions about detour focus on the scene that the destination is opposite to the origin. On the other hand, the destination is on the other side of the street diagonally and there are several crossing facilities located between them. In this scene, pedestrians would consider whether cross first or walk first. For example, the origin is in zone 1 and the destination is in zone 6 (as shown in Figure 1), there are two choices: to cross first via

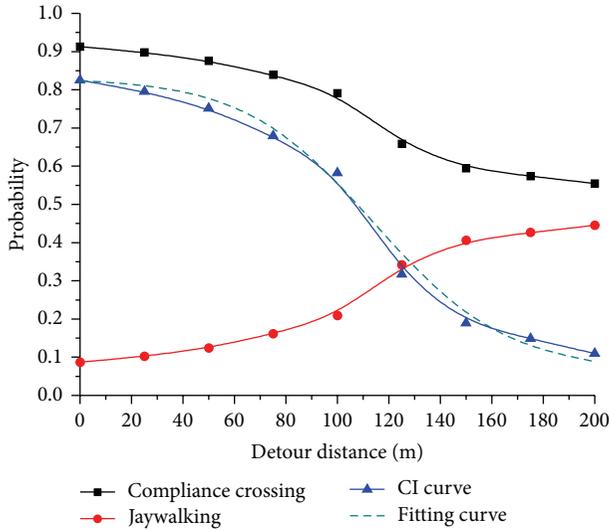


FIGURE 5: Illustration of compliance index.

the crosswalk or walk first then cross via the overpass. On the basis of the survey, the probabilities of the above questions are given by the following.

Cross first and walk later:  $P_c = 0.632$ .

Walk first and cross later:  $P_w = 0.368$ .

This means pedestrians are inclined to cross earlier. The result can be explained by the fact that people may miss a proper crossing location if they keep going. Especially for people who are not familiar with the locations of crossing facilities, they are willing to cross at first when they get a cross facility. Yannis et al. [15] obtained a similar result and gave a linear formula of probability distribution to describe the cross first behaviour. In the research of Yannis et al., a pedestrian's trip may include several blocks, so people may have diverse choices such as midblock or intersection. But it is not necessary for pedestrians to walk so long distance in most real conditions. While walking along the trip, pedestrians who are familiar with the traffic environment can choose the cross locations they prefer. However, the strangers are inclined to use the first facility they get. This phenomenon gives a message that crossing facilities should be near the generating/attracting source of walk trip.

## 5. Applications and Engineering Improvements

According to the analysis of pedestrian behaviour, it is known that pedestrians' subjective willingness can determine the final choice of crossing location. The influential factors lead to diverse choices in the same circumstance: some pedestrians prefer the crossing facilities, but some cross illegally. This is because pedestrians themselves are very appropriate to identify whether the traffic environments are desirable. Reasonable designing and operation of crossing facilities which match pedestrians' characteristics can conduct the pedestrians' behaviour and increase the use of pedestrian facilities.

So the requirements in designing of crossing facilities are to satisfy pedestrian needs and improve pedestrian safety and comfort. Here, some engineering improvements are discussed based on pedestrians' behavioural characteristics in order to make contribution to the traffic environments.

- Grade-separated crossing facilities are recommended for pedestrian safety and traffic performance. Though these facilities are built under severe construction conditions, pedestrian overpass or underpass is the best choice, especially in business quarters and residential areas.
- In the sight of pedestrian, the shorter interval distance is more convenient. Considering the requirements of performance and fairness, the commendatory interval distance between two crossing facilities is 300 meters (400 meters at the most). This is because a shorter interval will increase traffic delay and decrease the utility rate of facilities. On the contrary, the long interval may increase pedestrian violation.
- The consideration about the location and type of crossing facilities is very important. Pedestrians' behavioural characteristics should be considered seriously, as well as the land use around the street. The main generating/attracting sources of walk trip (e.g., school and shopping mall) should be concerned specially. In other words, crossing facilities mainly serve such trip sources.
- In response to the situation where pedestrian needs to detour, the strategy is the nearer the better. If the detour distance is inevitable, such distance is recommended to be less than 150 meters. This scheme aims to reduce the possibility of jaywalking and it is important for the area with large pedestrian flow.
- If there are trip generating/attracting sources at both side of the street diagonally, pedestrians will face the alternatives whether cross first or walk first. According to above discussion, pedestrians are inclined to cross first and walk later. Therefore it will be better to set up two facilities for each source if the interval is more than 200 meters. On the other hand, if the interval is shorter than 200 meters, it will be better to design a facility in the middle of them.

Finally, to strengthen the awareness of traffic safety is the most basic and the most efficient measure to minimize pedestrian violation.

## 6. Conclusions

This paper examines the pedestrians' preferences of crossing locations and the influential factors in making a decision to cross a street. Information is obtained through questionnaire survey and field observation in Beijing. A multinomial logit model is used to describe the crossing behaviour under various traffic conditions. And how the influential factors have impact on the crossing action is revealed by sensitivity analysis. The following conclusions are drawn based upon the results.

- (a) There is a relatively good agreement between the information of questionnaire and pedestrian actual behaviour in terms of their crossing choices. The MNL model can describe crossing behaviour preferably.
- (b) The most influential subjective factor in making a decision to cross at a designated crossing location is the action principle (variable  $X_{n3}$ ). The most influential factor in external environment is the detour distance (variable  $X_{n9}$ ).
- (c) A reasonable designing and a proper operation of crossing facilities can conduct crossing behaviour and increase the utilization rate of facilities. Pedestrians themselves are the most appropriate group to identify the rationality and serviceability of pedestrian environment.
- (d) The results indicate that pedestrians prefer overpass/underpass, and most pedestrians consider safety the first importance. In a view of cost and construction conditions, signalized crosswalk is recommended to help minimize the pedestrian-vehicle conflicts.

The findings from this paper are expected to help understand pedestrians' behaviour at crossing locations well. Though pedestrians' preferences and the reasons behind the choice are explained by a discrete choice model, there are still some insufficiencies in the work. For example, some potential influential factors may be neglected or the factor's influencing degree is weakened so that the model cannot reflect the effects of variables comprehensively. Additionally, aggregate results for crossing behaviour with various OD pairs and the reason for pedestrian violation should be considered further.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was supported in part by the National Nature Science Foundation of China under Grants 51378062 and 71301010 and the Introducing Talents of Discipline to Universities under Grant B12022.

## References

- [1] M. Huan, X. B. Yang, and B. Jia, "Crossing reliability of electric bike riders at urban intersections," *Mathematical Problems in Engineering*, vol. 2013, Article ID 108636, 8 pages, 2013.
- [2] J. J. Fruin, *Pedestrian Planning and Design*, Metropolitan Association of Urban Designers and Environmental Planners, New York, NY, USA, 1971.
- [3] S. T. Mohammed, "Evaluation of pedestrian speed in Jordan with investigation of some contributing factors," *Journal of Safety Research*, vol. 32, no. 2, pp. 229–236, 2001.
- [4] W. H. K. Lam, J. Y. S. Lee, and C. Y. Cheung, "A study of the bi-directional pedestrian flow characteristics at Hong Kong signalized crosswalk facilities," *Transportation*, vol. 29, no. 2, pp. 169–192, 2002.
- [5] V. P. Sisiopiku and D. Akin, "Pedestrian behaviors at and perceptions towards various pedestrian facilities: an examination based on observation and survey data," *Transportation Research F: Traffic Psychology and Behaviour*, vol. 6, no. 4, pp. 249–274, 2003.
- [6] S. Sarkar, "Evaluation of safety for pedestrians at macro- and microlevels in urban areas," *Transportation Research Record*, no. 1502, pp. 105–118, 1995.
- [7] J. Y. S. Lee, P. K. Goh, and W. H. K. Lam, "New level-of-service standard for signalized crosswalks with bi-directional pedestrian flows," *Journal of Transportation Engineering*, vol. 131, no. 12, pp. 957–960, 2005.
- [8] R. Elvik, M. W. J. Sørensen, and T.-O. Nævestad, "Factors influencing safety in a sample of marked pedestrian crossings selected for safety inspections in the city of Oslo," *Accident Analysis and Prevention*, vol. 59, pp. 64–70, 2013.
- [9] K. Shriver, "Influence of environmental design on pedestrian travel behavior in four Austin neighborhoods," *Transportation Research Record*, no. 1578, pp. 64–75, 1997.
- [10] B. Li, "A model of pedestrians' intended waiting times for street crossings at signalized intersections," *Transportation Research B: Methodological*, vol. 51, pp. 17–28, 2013.
- [11] P. Lambrianidou, S. Basbas, and I. Politis, "Can pedestrians' crossing countdown signal timers promote green and safe mobility?" *Sustainable Cities and Society*, vol. 6, pp. 33–39, 2013.
- [12] H. Guo, W. Wang, W. Guo, X. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [13] H. Guo, Z. Gao, X. Yang, and X. Jiang, "Modeling pedestrian violation behavior at signalized crosswalks in China: a hazards-based duration approach," *Traffic Injury Prevention*, vol. 12, no. 1, pp. 96–103, 2011.
- [14] X. Chu, M. Guttenplan, and M. R. Baltes, "Why people cross where they do: the role of street environment," *Transportation Research Record*, no. 1878, pp. 3–10, 2003.
- [15] G. Yannis, J. Golias, and E. Papadimitriou, "Modeling crossing behavior and accident risk of pedestrians," *Journal of Transportation Engineering*, vol. 133, no. 11, pp. 634–644, 2007.
- [16] H. Nassiri and Y. Sajed, "Using a logit model to predict pedestrian crossing behaviour based upon vehicle speed and headway on multi-lane street," in *Proceedings of the 88th Annual Meeting of the Transportation Research Board*, Washington DC, USA, 2009.
- [17] M. Ben-Akiva and L. Steven, *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Mass, USA, 1985.

## Research Article

# Application of the Expanded Theory of Planned Behavior in Intercity Travel Behavior

Jing Peng,<sup>1,2</sup> Juan Zhi-cai,<sup>1</sup> and Gao Lin-jie<sup>1</sup>

<sup>1</sup> Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai 212013, China

<sup>2</sup> School of Automobile and Traffic Engineering, Jiangsu University, Zhenjiang, Jiangsu 200052, China

Correspondence should be addressed to Jing Peng; jingpeng@ujs.edu.cn

Received 31 October 2013; Accepted 18 January 2014; Published 5 March 2014

Academic Editor: Huimin Niu

Copyright © 2014 Jing Peng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Congestion in intercity corridors of metropolitan area has been increasing steadily. To alleviate congestion, many major investment projects, such as the high speed railway projects, were proposed by agency. To evaluate the adequacy and efficiency of these projects, the intercity travel behavior should be analyzed in metropolitan area. The paper constructed a Multiple Indicators and Multiple Causes (MIMIC) model according to an expanded theory of planned behavior (TPB) to study the travel behavior of choosing from the choice set of the traditional train, the high speed railway and the coach by demographic and psychological factors. Through empirical data collection and analysis, we found that demographic factors of travelers indeed positively engender the latent variables in MIMIC, and descriptive norm and habit had direct or indirect significant effect on travel behavior and intention. On the basis of the effect of psychological constructors of the expanded TPB on the intercity travel behavior and differentiation of traveler's demographic characteristics, the agency can make reasonable policies and proper information for the intercity transportation. The results will support the basic theory of optimizing the transportation system in metropolitan area. Implications for researchers and suggestions for future research are also addressed in this study.

## 1. Introduction

Congestion in intercity corridors of metropolitan area has been increasing steadily, which has raised serious concerns for its adverse impacts on regional economic development, national productivity and competitiveness, and environmental quality [1]. To alleviate congestion, many major investment projects, such as the Electric Multiple Unit (EMU) and the high speed railway (HSR) projects, were proposed by agency. To evaluate the adequacy and efficiency of these projects, public agencies need analyses of the intercity travel demand in metropolitan area due to the limited allocated financial resources. Meanwhile, intercity traveler carriers welcome reliable forecasts of intercity demand so that they can be more responsive to their patronage and to remain competitive. Therefore, intercity travel behavior research in metropolitan area is needed to estimate and evaluate expected policy impacts. Transportation agencies often focus on influencing travel behavior by changing the physical system. The possibility of actively influencing traveler preferences through

the psychological factors opens a whole new set of options that have been largely overlooked in the past [2]. Combining both perspectives enables agencies to position transportation policy within the broad context of sustainable metropolitan management.

This study compares psychological predictors of the intention and behavior to use three intercity travel modes in metropolitan area of Yangtze River Delta in China: the traditional train, the HSR, and the coach, and examines the effect that habit operates as moderator of intention-behavior relationship. In previous studies, various theoretical perspectives have been employed to understand factors important for choosing travel modes. The most widely applied model of cognitive determinants of choosing travel modes is the theory of planned behavior (TPB), which suggests that behavior is most closely determined by an intention to act [3]. Intentions are based on a combination of attitude toward the behavior, subjective norm (SN), and perceived behavioral control (PBC). Intention has a direct effect on behavior, and, under some circumstances, the same applies to perceived behavioral

control. The theory recognizes the importance of background factors, such as personality, emotions, education, age, gender, and past experience; although if they affect behavior, it would be via beliefs.

The TPBs sufficiency assumption is invalid; in other words, intention may be determined not only by attitude, subjective norm, and perceived behavioral control but also by more additional variables [4]. In this study, the original TPBs predictive validity was tested in the behavior of using the three intercity travel modes in Yangtze River Delta area. Furthermore, descriptive norm and habit as additional predictors of intention and behavior in TPB were examined, because the two variables have been provided enough evidences to satisfy Ajzen's criteria of adding predictors in other behavior domains [4–7]. The paper is organized as follows. The next section discusses the two additional variables expanding the TPB. Section 3 proposed hypothesis with respect to relationship among variables of the expanded TPB. Section 4 describes the data collection process and data used in the research. Section 5 presents the MIMIC model and the estimation results are given. The paper ends with a customary section of conclusions.

## 2. Descriptive Norm and Habit

According to Cialdini et al., a distinction should be made between subjective norms and descriptive norms (DN) [8]. The former refers to beliefs about what are and what are not approved ways of conduct, what one ought to do, while descriptive norms are what are typical and normal behaviors. With regard to mode choice, studies have found that descriptive norm is a significant predictor of using the car and the bus whereas subjective norm significantly predicted the intention to use these modes [9, 10], although Thøgersen found that only subjective norm was a significant predictor of using public transport [11]. Several explanations have been given as to why descriptive norms are sometimes more important than subjective norms. One methodological explanation is that since descriptive norms often display lower means and larger variability compared to subjective norms, the risk of reduced variability is greater for subjective norms [12]. Other explanations are that the distinction is real and not only a methodological artifact.

The habit approach is consistent with the TPB in that it suggests that, faced with a new or unfamiliar choice situation, a traveler will deliberate and form an intention to choose the most attractive goal-directed option, which will inform subsequent behavior [13]. The more frequently a behavior has been performed in a stable context, the more it is said to habituate and come under the direct control of external stimulus cues at the expense of intentions [14]. Habits are most clearly revealed where habitual and intentional tendencies diverge, because in such situations behavioral outcomes will correspond with habits but not intentions. The conflicted relationship between habit and intention in the extant transport literature potentially overlooks that, in the absence of modifications of the decisional environment, habits are

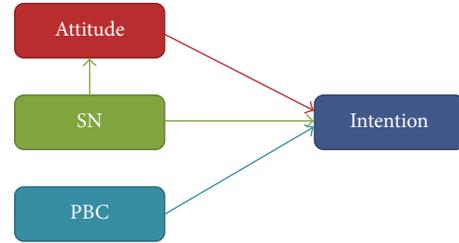


FIGURE 1: Impact of subjective norm on attitude.

likely to correspond with intentions, having developed from frequently enacted intentions.

## 3. Hypothesis

Even though several studies have examined psychological predictors of travel mode choice, the focus has often been on examining a single travel mode making a comparison between different travel modes unfeasible. Moreover, in the context of mode choice, few studies have treated subjective and descriptive norms as two separate constructs and examined the role of habit in intention-behavior relationship in the behavior domain of choosing intercity travel modes. Through testing the following hypotheses, we examine psychological predictors of the expanded TPB and relationships among them.

*3.1. The Relationship of Subjective Norms towards Attitudes.* Wu and Lin revealed that subjective norms can directly influence attitude [15]. Both have a significant relationship with each other. As the positive support received by individuals from other persons or organizations important to them becomes greater, their attitude also becomes more positive [16]. When the subjective norms of respondents are more positive, their attitudes also become more positive. Research of Yu into the behavior patterns of downloading MP3 shows that the subjective norm of users on downloading MP3 positively influences their attitude. In view of these, we present the first hypothesis as follows: there exists a significant relationship between subjective norm and participation attitude in intercity travel mode choice behavior, as shown in Figure 1.

*3.2. The Role of Descriptive Norms in Expanded TPB.* Ravis used meta-analysis to find a medium to strong average correlation between descriptive norms and intention and, more importantly, showed a significant improvement in the predictive validity of the TPB when descriptive norm was included as an additional predictor [7]. Therefore the second hypothesis is as follows: the descriptive norms have positive significant impact on intention to choose intercity travel mode. Accounting for the descriptive norms and subjective norms being parts of social norms in psychological theory, the descriptive norms' significant influences on attitudes towards intercity travel modes became the third hypothesis of the paper. As shown in Figure 2.

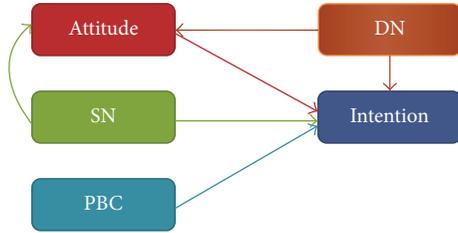


FIGURE 2: Impact of descriptive norm in TPB.

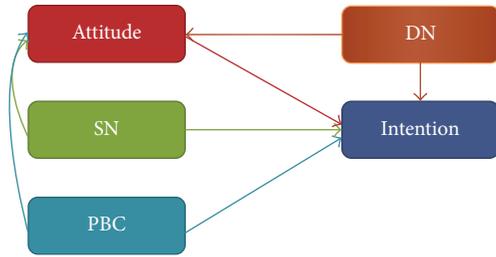


FIGURE 3: Impact of PBC on attitude.

3.3. *The Relationship of Perceived Behavioral Control towards Attitudes.* Ajzen’s original TPB did not include the relationship that PBC has influence on attitude. However, attitude can be an intervening variable of the subjective norm when influencing behavioral intention. Thus, in the causal model constructed by Yu specifying the behavioral tendencies of Taiwanese tourists in Kinmen [16], attitude was made an intervening variable. Results of this study show attitude as an intervening variable in the effect of perceived behavioral control towards behavior intention. Tsai also proved the impact of perceived behavioral control on attitude by analysis of canonical correlation [17]. From this, it can be said that perceived behavioral control has a positive effect on attitude. Taking these into account, the paper gives its forth hypothesis as follows: there exists a significant relationship between perceived control behavior and participation attitude. As shown in Figure 3.

3.4. *The Role of Habit in the Expand TPB.* Habit has a great impact on individuals’ choice behavior. Several empirical studies have found that habit included in TPB showed significant relationships with all the other original variables [6, 18, 19]. The fifth hypothesis proposed is that habit has significant impact on attitude, subjective norms, descriptive norms, perceived behavior control, and intention to choose intercity travel modes. According to Gardner’s research that habit will moderate the effect of intention on behavior [9]: where habit is weak, intention will predict behavior, but where habit is strong, intention will have a weak effect on behavior; Figure 4 illustrates the sixth hypothesis: habit has moderate effect between intention and behavior on intercity travel mode choice behavior.

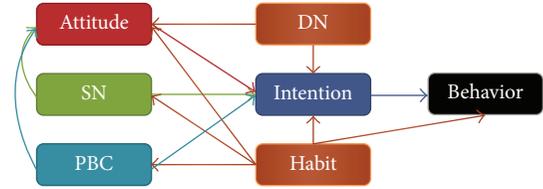


FIGURE 4: Impact of habit in TPB.

TABLE 1: Socioeconomic status variables.

Categories	Value	Frequency (%)	Variables
Gender	Male	58.41	Male
	Female	41.59	
Age	16~25	27.11	Young
	25~54	54.61	
	55+	18.28	Old
Employment	Working	58.34	Working
	Student	26.85	Student
	Other	14.81	
Income	Unites: RMB	—	Income
Education	(1) Primary;	—	Edu
	(2) junior;		
	(3) senior;		
	(4) technical secondary school;		
	(5) college;		
	(6) undergraduate;		
	(7) master;		
	(8) Ph.D.		

#### 4. Data

The study was conducted in Zhenjiang in the east of China. Zhenjiang is one of the most important transport hubs in Yangtze River Delta area. In 2012, a survey was carried out by face-to-face interviews with the travelers in railway station and long distance bus station. A questionnaire was sent to a randomly selected sample of 3695 passengers in the station and the response rate was 93%. The number of completed questionnaires was 3436, of which 3248 individuals are included as the study sample after removing those who travel outside Yangtze River Delta area. Demographic characteristics concerning gender, age, education, occupation, income, and access to various travel modes were assessed in the questionnaire at first, as shown in Table 1.

More men than women answered the questionnaire, the mean age was 26, and more than 56% had university or college education. In the questionnaire, the participants described the trip’s purpose, how long it was, and what travel mode was often used. Forty-two percent made a work trip, 21% a trip to visit relatives and friends, 5% to study, 3% to the doctor, 13% to leisure activities, 3% to service and shopping activities, 8% to go home, and 5% had other trip purposes. The most command travel mode was the HSR, chosen by 36.98% of the participants, while 32.11% used the coach, and 30.91% used the traditional train.

Following the description of the typical set of socioeconomic variables, the participants evaluated the use of the traditional train, HSR, and coach. The survey contains

29 psychometric indicators with respect to relevant TPB variables.

Attitudes towards the three intercity travel modes were accessed by a combination of behavioral beliefs and outcome evaluation. Initially, the respondents were solicited to rate the consequences of using the three intercity travel modes, 7 different behavioral beliefs on a five-point scale (1 = strongly disagree, 5 = strongly agree). The behavioral beliefs included fitness level, feeling of free and relax, expenditure for the ticket, risk of being in a traffic accident, quickly, convenient for transfer and buying the ticket. Subsequently, the importance of each of the consequences was evaluated on a five-point scale (1 = not at all important to me, 5 = very important to me) to give the outcome evaluations. Before combining the behavioral beliefs and outcome evaluations into measures of attitudes toward the three modes, positive consequences were recoded to make sure those higher values on all behavioral beliefs indicated a more positive belief. Each behavioral belief was multiplied by the respective outcome evaluations. The products were summarized and divided by the number of items resulting in a scale from 0.5 to 12.5 where a higher value signified a more positive attitude. Principle components analysis identified just one component accounting for 72% of the variance (eigenvalue of 2.75). Cronbach's alpha ( $\alpha$ ) was 0.75, indicating strong inner consistency.

The two types of social norms were measured by two indicators for both subjective norms and descriptive norms. For subjective norms the items "My best friends consider using the traditional train/using the HSR/using the coach to be ..." and "My family/relatives consider using the traditional train/using the HSR/using the coach to be ..." were assessed on a five-point scale ranging from completely unacceptable to completely acceptable. The items were recoded so that a high value indicated a stronger subjective norm.

Principle components analysis identified a single coherent component, accounting for 76% of the variance (eigenvalue of 1.52). Cronbach's alpha was lower than that for intention ( $\alpha = 0.68$ ) but still indicates reasonable internal consistency. Nunnally suggested that  $\alpha = 0.70$  represents strong inner consistency [20], but Cortina urges researchers to consider the number of items used—a moderate alpha with small number of items may well represent better internal consistency than a larger alpha with a larger number of items [21, 22]. Ajzen suggests that a requirement for high internal consistency for belief based measures is not necessary, given that it is the aggregate of differing beliefs that forms an attitude [14]. The principle components analysis showing that the aggregated variable forms a unitary component is an important justification for aggregation.

Descriptive norms were measured by the items "My closest friends will themselves use the traditional train/use the HSR/use the coach" and "My family/relative will themselves use the traditional train/use the HSR/use the coach" and rated a five-point scale from strongly disagree to strongly agree. After recoding the items, a high value indicated a strong descriptive norm. Again, principle components analysis revealed one component accounting for 72% of the variance (eigenvalue = 1.44). Cronbach's alpha indicated

moderate internal consistency, again in line with expectations for such a belief based aggregate,  $\alpha = 0.61$ .

Direct measures of perceived behavioral control were used, including three items for each mode: (i) "It's mainly up to me whether I choose the intercity travel mode or not"; (ii) "To use the travel mode on my ordinary trip is difficult"; (iii) "It will make me feel trouble to choose the travel mode". All the three items were evaluated on five-point scale ranging from strongly disagree to strongly agree. Subsequently, the items were recoded so that a higher value indicated a higher perceived behavioral control. These items formed one component in a principle components analysis accounting for 65% of the variance (eigenvalue = 1.95). Cronbach's alpha ( $\alpha$ ) was 0.73, indicating strong inner consistency.

Behavior intention was assessed separately for different travel mode by a mean of three items: (i) "It is likely that I will choose the intercity travel mode in the future"; (ii) "I would expect to use the intercity travel mode in the next time"; (iii) "Within the next coming one month I have the intention to use the travel mode". All items were evaluated on a five-point scale (1 = completely impossible, 5 = completely possible). After recoding the items, a higher value signified a stronger intention to use that particular intercity travel mode. Principle components analysis identified just one component accounting for 65% of the variance (eigenvalue of 1.94). Cronbach's alpha ( $\alpha$ ) was 0.72, indicating strong inner consistency.

In addition to assessing the predictors in the TPB, the questionnaire also inquired into the respondents' behavior by asking them to indicate how often (1 = always, 5 = never) and frequently (1 = very low, 5 = very high) they had used each intercity travel mode. Principle components analysis identified just one component accounting for 77% of the variance (eigenvalue of 1.53). Cronbach's alpha ( $\alpha$ ) was 0.69, indicating reasonably inner consistency.

Habit was measured using a ten-item version of Verplanken and Orbell's Self-Report Habit Index (SRHI) [23]. Each item related to "Choosing the travel mode on the intercity trip" (e.g., "Choosing the traditional train on the inter-city trip is something I do automatically," "Choosing the traditional train on the inter-city trip is something I do without having to consciously remember") and was measured on a five-point scale (1 = strongly disagree, 5 = strongly agree;  $\alpha = 0.80$ ).

The indexes of reliability and validity for the predictors of the TPB were listed in Table 2.

## 5. Model and Estimation

In order to examine the interrelationships among the latent variables of TPB and between them and the socioeconomic status variables, a Multiple Indicators and Multiple Causes (MIMIC) model is estimated. In terms of the multivariate regression of the indicators on the causes, the model implies restrictions of two types: (i) the regression coefficient matrix has ranked one; (ii) the residual variance-covariance matrix satisfies a factor analysis model with one common factor.

TABLE 2: Measurements of reliability and validity.

Variables	Percentage of one component accounting for the variance	Eigenvalue	Cronbach's alpha ( $\alpha$ )
Attitude	72%	2.75	0.75
SN	76%	1.52	0.68
DN	72%	1.44	0.61
PBC	65%	1.95	0.73
Intention	65%	1.94	0.72
Behavior	77%	1.53	0.69
Habit	71%	2.73	0.80

The MIMIC model is a special form of structural equation modeling (SEM) in fact. The specification of the model is as follows:

$$\boldsymbol{\eta} = \Gamma \mathbf{x} + \zeta \quad (1)$$

$$\mathbf{y} = \Lambda \boldsymbol{\eta} + \varepsilon, \quad (2)$$

where (1) is structural equation and (2) is measurement equation. The latent variable vector  $\boldsymbol{\eta}$  is linearly determined, subject to disturbances  $\zeta$ , by vector of observable exogenous causes  $\mathbf{x}$ . The latent variable and disturbance  $\varepsilon$  determine the vector of observable endogenous indicators  $\mathbf{y}$  linearly.  $\Gamma$  and  $\Lambda$  are matrices of unknown parameters to be estimated. The operational implications of the model appear when we solve for the reduced-form relation connecting the observables:

$$\mathbf{y} = \Lambda (\Gamma \mathbf{x} + \zeta) + \varepsilon = \Pi \mathbf{x} + \mathbf{v}, \quad (3)$$

where the reduced-form coefficient matrix is

$$\Pi = \Lambda \Gamma \quad (4)$$

and the reduced-form disturbance vector is

$$\mathbf{v} = \Lambda \zeta + \varepsilon. \quad (5)$$

Estimation of a structural equation latent variable model minimizes the difference between the sample covariance matrix,  $\mathbf{S}$ , and the covariance matrix  $\boldsymbol{\Sigma}$ . The elements of  $\boldsymbol{\Sigma}$  are hypothesized to be a function of the parameter vector  $\boldsymbol{\theta}$  so that  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ . The parameters are estimated so that the discrepancy between  $\mathbf{S}$  and the implied covariance matrix  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$  is minimal. The discrepancy function,  $F = F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ , measures the discrepancy between  $\mathbf{S}$  and  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  evaluated at  $\hat{\boldsymbol{\theta}}$ .  $F_{\min}$  is the minimum value of the discrepancy function and equals zero only if  $\mathbf{S} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ . An indication of model fit is, therefore, given by the closeness of the  $F_{\min}$  to zero, supposing that the disturbances are all mutually independent. For convenience, the expectation of all variables is zero:

$$E(\zeta \varepsilon') = 0, \quad E(\zeta^2) = \sigma^2, \quad E(\varepsilon \varepsilon') = \Theta, \quad (6)$$

where  $\Theta$  is the diagonal matrix with  $\theta$ , the vector of standard deviations of the  $\varepsilon$ 's, displayed on its diagonal. The covariance matrix can be computed by

$$\sum(\hat{\boldsymbol{\theta}}) = E(\mathbf{v} \mathbf{v}') = \sigma^2 \Lambda \Lambda' + \Theta. \quad (7)$$

TABLE 3: Goodness-of-fit statistics for MIMIC model.

	Traditional train	High speed railway	Coach
$\chi^2$	772.865	787.114	834.648
df	235	235	235
$\chi^2/\text{df}$	3.289	3.349	3.552
RMSEA	0.048	0.044	0.049
CFI	0.890	0.903	0.880
TLI	0.858	0.875	0.845
SRMR	0.043	0.046	0.044

The multiple indicator part of the MIMIC model is a confirmatory factor analytical model specified. The multiple cause part of the model is given by

$$\begin{aligned} \eta_{li} = & \gamma_{11} \text{male}_i + \gamma_{12} \text{young}_i + \gamma_{13} \text{old}_i \\ & + \gamma_{14} \text{working}_i + \gamma_{15} \text{student}_i + \gamma_{16} \text{income}_i \\ & + \gamma_{17} \text{edu}_i + \zeta_i \end{aligned} \quad (8)$$

$l = \text{Attitude, SN, DN, PBC, Intention, Habit.}$

Figure 5 illustrates the structure of the MIMIC model. Structural equation and measurement equation are alleviated to SE and ME in Figure 5.

As illustrated in Figure 5, the MIMIC model includes demographic characteristics of travelers, the latent variables that construct the expanded TPB, and endogenous observed indicators. Specifically this model hypothesizes that the socioeconomic variables influence all latent variables of TPB, which are also explained by indicators from questionnaires for respondents. Figure 6 specifies the hypothesized relationships among the latent factors, where ellipses represent unobservable variables and rectangles observable indicators. Dashed arrows represent measurement equations while solid arrows represent the structural equations. The latent variable model describes the relationships between the latent variables and their indicators and causes.

The MIMIC model simultaneously estimates the measurement equations relating each factor to its indicators, and the structural equations specify the relationships among latent factors and between them and socioeconomic status variables. The estimation of the MIMIC model was conducted in STATA 12. Table 3 summarizes the overall goodness-of-fit statistics.

Most interpreters of the root mean squared error of approximation (RMSEA) test label the fit close if the lower bound of the 90% CI is below 0.05 and label the fit poor if the upper bound is above 0.10. CFI and TLI are two indices such that a value close to 1 indicates a good fit. CFI stands for comparative fit index. TLI stands for Tucker-Lewis index and is also known as the nonnormed fit index. A perfect fit corresponds to a standardized root mean squared residual (SRMR) of 0. A good fit is a small value, considered by some to be limited to 0.08. Though CFI is 0.880 slightly below 0.9, RMSEA and SRMR are below 0.05 and in particular the full 90% confidence interval 0.027~0.030 falls below 0.05 so the overall data fit is acceptable; that is, the model cannot reject

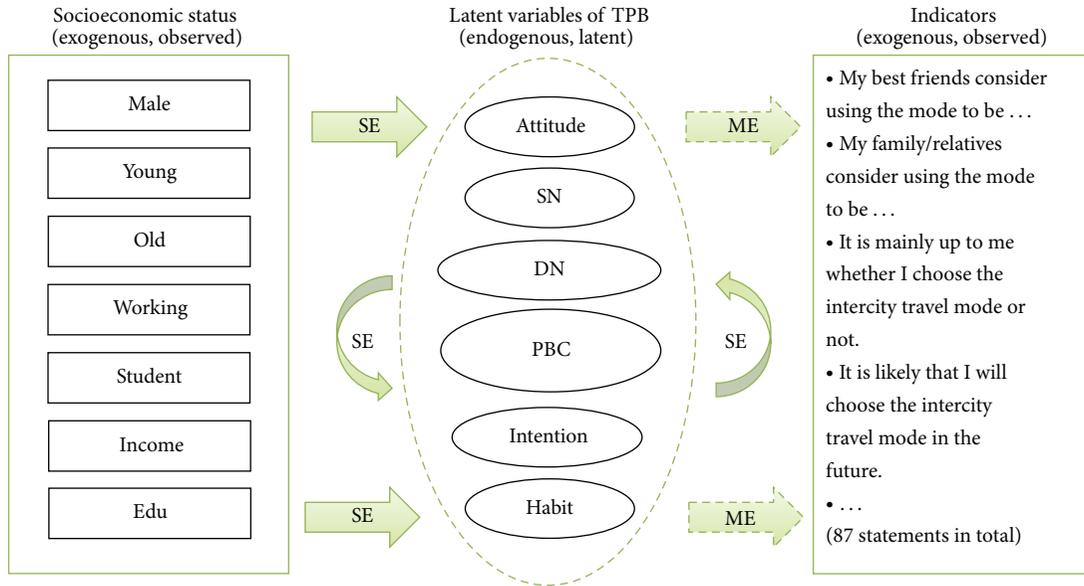


FIGURE 5: Multiple Indicators and Multiple Causes (MIMIC) model.

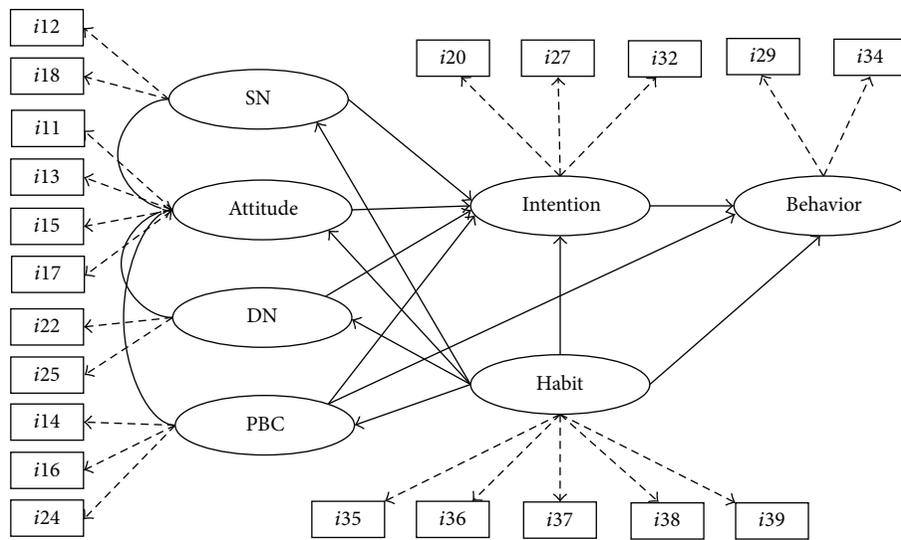


FIGURE 6: Detailed path analysis diagram.

the hypothesis of the relationships among the latent factors and between them and demographic variables specified in Figures 5 and 6.

## 6. Results

Based on the results from the MIMIC model, we can examine the relationships between the demographic characteristics variables and the latent variables in TPB and within them.

### 6.1. Relationships between the Demographic Variables and the Latent Factors.

The relationships between the demographic variables and the latent factors are summarized as the regression coefficients shown in Table 4. The number in parenthesis is  $t$  statistics.

As shown in Table 4, young and education have significant impact on some of the latent variables in MIMIC models using all the three intercity travel modes. However, career and income play a significant role in the model using coach.

In the MIMIC model of the traditional train, male and old are significantly associated with liking to habit. Education has a significant negative impact on habit. People with less years of education are more accustomed to use traditional train as intercity travel mode. Being working or income is not associated with the latent variables of TPB. People of male and being students have significant influence on PBC, which

TABLE 4: Impact of demographical variables on the latent factors.

	Male	Young	Old	Working	Student	Income	Edu
Traditional train							
SN	0.022 (0.56)	-0.074 (-1.41)	-0.040 (-0.73)	-0.060 (-1.11)	-0.010 (-0.13)	-0.019 (-0.97)	0.018 (1.36)
DN	0.005 (0.15)	<b>-0.141** (-3.26)</b>	0.018 (0.40)	-0.007 (-0.15)	<b>0.150* (2.25)</b>	0.027 (1.64)	0.014 (1.32)
Attitude	-0.017 (-0.48)	-0.042 (-0.72)	-0.019 (-0.41)	0.078 (1.67)	0.066 (0.86)	-0.009 (-0.53)	-0.005 (-0.39)
PBC	<b>0.094* (2.45)</b>	-0.063 (-1.25)	-0.048 (-0.93)	0.041 (0.79)	<b>0.164* (2.11)</b>	0.017 (0.89)	0.024 (1.84)
Habit	<b>0.103* (2.53)</b>	0.031 (0.57)	<b>0.164** (2.96)</b>	-0.078 (-1.43)	-1.30 (-1.60)	0.013 (0.65)	<b>-0.056** (-4.12)</b>
High speed railway							
SN	0.023 (1.03)	-0.024 (-0.66)	0.024 (0.73)	0.009 (0.28)	0.015 (0.32)	<b>-0.022* (-2.16)</b>	<b>0.024** (2.67)</b>
DN	-0.027 (-1.01)	-0.011 (-0.33)	0.008 (0.21)	0.032 (0.79)	-0.034 (-0.64)	-0.022 (-1.79)	0.014 (1.36)
Attitude	-0.20 (-1.28)	<b>-0.094** (-2.91)</b>	<b>0.068* (2.00)</b>	-0.06 (-0.26)	<b>0.068* (2.04)</b>	-0.08 (-0.99)	0.001 (0.17)
PBC	0.005 (0.16)	0.033 (0.87)	-0.025 (-0.57)	0.086 (1.89)	<b>0.142* (2.34)</b>	0.018 (1.29)	0.018 (1.63)
Habit	<b>-0.082* (-2.39)</b>	-0.034 (-0.78)	-0.032 (-0.62)	-0.042 (-0.80)	0.039 (0.56)	<b>0.040* (2.53)</b>	<b>0.036** (2.81)</b>
Coach							
SN	-0.039 (-1.10)	0.037 (0.77)	-0.008 (-0.15)	-0.089 (-1.57)	-0.035 (-0.50)	0.001 (0.06)	0.022 (1.69)
DN	0.016 (0.48)	<b>-0.107* (-2.30)</b>	0.033 (0.63)	-0.038 (-0.69)	-0.051 (-0.73)	-0.011 (-0.63)	0.007 (0.55)
Attitude	0.012 (0.32)	-0.048 (-0.89)	-1.04 (-1.70)	0.060 (0.94)	-0.006 (-0.08)	0.002 (0.10)	<b>-0.033* (-2.26)</b>
PBC	-0.002 (-0.05)	0.025 (0.50)	-0.022 (-0.38)	<b>-0.147* (-2.47)</b>	-0.087 (-1.16)	0.015 (0.79)	0.003 (0.27)
Habit	<b>0.099* (2.36)</b>	-0.028 (-0.49)	0.023 (0.36)	0.017 (0.26)	-0.150 (-1.75)	<b>-0.053* (-2.48)</b>	-0.004 (-0.28)

\*  $P < 0.05$ ; \*\*  $P < 0.01$ .

implies that man and students may have more patience than female and workers for the bad schedules of traditional train. Being young and students are significantly associated with descriptive norms, while students think that their family or relatives should prefer the traditional train.

In the MIMIC model of the high speed railway, gender, income, and education of all the travelers' demographical characteristics have significant influence on habit. This result seems logical considering the fact that female to a great extent than men are accustomed to use high speed railway as intercity travel mode. Being student has positive impact on perceived behavioral control of high speed railway, which may imply that students are open to new intercity travel mode. The pairwise correlation coefficients between income and being young display that the two explanted exogenous observed variables have statistical significant correlation at the individual 1% level. Although income has not direct significant influence on attitude of high speed railway, we can see from the relationship of being young and attitude that higher incomes are coupled with stronger preferences for habit, potentially reflecting the fact that the opportunity cost of time losses is higher at higher incomes.

In the MIMIC model of coach, respondents with male and lower incomes are used to travel with coach as intercity mode. Considering the indicators used to construct perceived behavioral control of coach shows that respondents with having work do not think that they can be fitted with the schedule of coach. In addition, negative significant impact of education on attitude of coach shows that respondents with lower education have stronger preferences for coach. Finally, considering the indicators used to construct descriptive norms it seems natural that respondents with being young

would expand their negative feelings with relation to coach to the important person for them.

Overall, the demographical characteristics of intercity travelers have different impact on the endogenous latent variables of TPB. Gender has significant influence on habit of all the three intercity travel modes, which shows that male is used to using traditional train and coach and female has stronger preferences for high speed railway. The negative significant impact of being young on descriptive norms of traditional train and coach shows that the younger is not satisfied with the perceptions of the two intercity travel modes, while the older respondents implicate that they are accustomed to using traditional train. Being student reflects the strong ability to adapt perceived behavior control of traditional train and high speed railway, while the workers express their to some extent distrust with respect to the ability of on time and convenience of coach. The respondents with higher incomes are habituated to high speed railway; the corresponding fact is that respondents with lower incomes are used to coach. The respondents' level of education shows negative impact on the latent factors of traditional train and coach; however, it has positive influence on habit and subjective norms of high speed railway, which signify that the respondents with high level of education ask for higher requirement of travel environment, timeline, and convenience.

6.2. *Relationships among the Latent Factors.* The significant interrelationships among the latent factors are summarized in Figures 7, 8, and 9, which shows the standardized coefficients ( $z$ -statistics in parenthesis) between latent factors in the path diagram. \* indicates  $P < 0.05$ ; \*\* indicate  $P < 0.01$ .

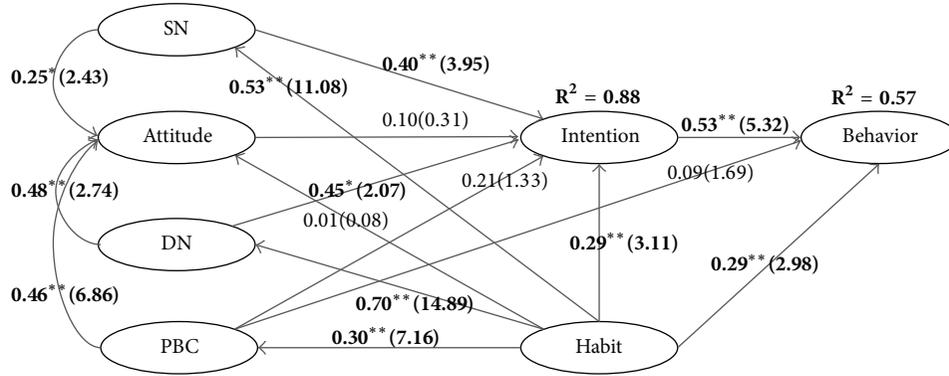


FIGURE 7: Standardized coefficients between the latent factors using traditional train.

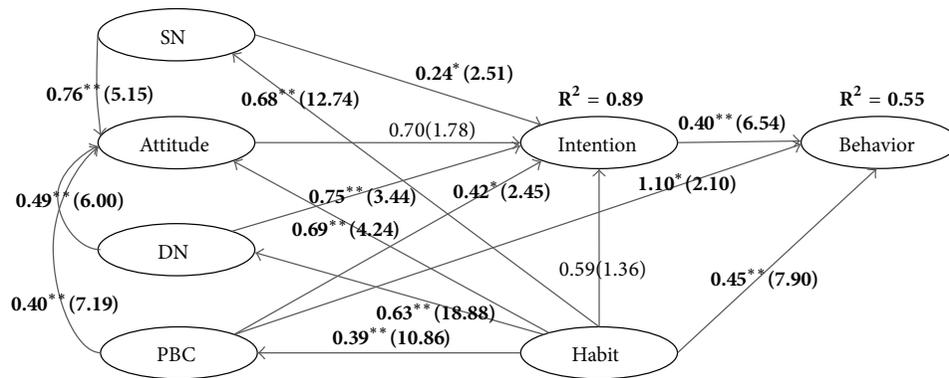


FIGURE 8: Standardized coefficients between the latent factors using HSR.

We can see from the three figures above that the MIMIC models with the expanded theory of planned behavior account for 77%~89% of the variance of intention to use the three intercity travel modes, which is consistent with Bamberg’s previous finds [24]. In Bamberg’s research the models with introduction of habit account for 77% and 80% of the variances in intention to choose the bus and car. The amounts of expanded variances in behavior of choice between intercity modes are 55%~61%, which of Bamberg are 51% and 66%. Although there are distinctions between the two researches, such as travel modes and latent factors, the amounts of expanded variances in behavior and intention to choose between modes to some extent imply that the expanded theory of planned behavior can be fitted with the intercity mode choice.

In the three MIMIC models, subjective norms, descriptive norms, and perceived behavioral control have significant impact on attitude at the individual 0.1% level, which support the first, third, and fourth hypothesis. The positive significant impact of habit on attitude in the mode of using high speed railway supports the fifth hypothesis in part. Descriptive norms have positive significant influence on intention to using the three intercity modes, which supports the second hypothesis. Except for the model of traditional train, descriptive norms have direct and indirect influence on choice behavior, as shown in Table 5. The number in parenthesis is *t* statistics. Hierarchical regression analyses are performed

TABLE 5: Direct and indirect impact of DN on intention and behavior.

	Traditional train	High speed railway	Coach
Direct impact on intention	0.45* (2.07)	0.75** (3.44)	0.28** (3.49)
Indirect impact on behavior	0.44 (1.86)	0.33* (1.97)	0.19** (3.87)

\**P* < 0.05; \*\**P* < 0.01.

and find that the explained variance increases significantly 4%~8% in all the models after the inclusion of descriptive norms, which verify that the influences of social norms on travelers are mutual feedback.

The relationship between habit and subjective norms, descriptive norms, perceived behavioral control, and behavior suggests that what take the travelers make decision among intercity mode is habit, which drives all the other latent factors. The standardized coefficients among habit, subjective norms, and descriptive norms are 53%~70% at the individual 0.1% level, which supports the fifth hypothesis in part. Habit has no significant impact on intention to using high speed railway, whereas it has that on intention to using traditional train and coach, which can be interpreted that high speed railway has not become the habitual mode for intercity

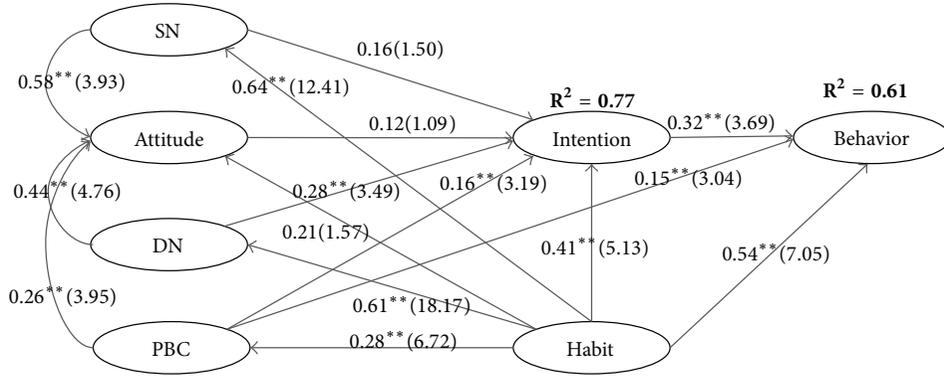


FIGURE 9: Standardized coefficients between the latent factors using coach.

TABLE 6: Habit as moderate between intention and behavior.

	Traditional train	High speed railway	Coach
Inclusion of habit	0.53** (5.32)	0.40** (6.54)	0.32** (3.69)
Exclusion of habit	0.73** (19.51)	0.68** (17.54)	0.73** (19.28)

\* $P < 0.05$ ; \*\* $P < 0.01$ .

travelers. Table 6 shows the standardized coefficients between intention and behavior to the exclusion of habit. The number in parenthesis is  $t$  statistics.

As illustrated in Table 6, the standardized coefficients between intention and behavior that arise in the MIMIC modes excluded the latent factor of habit. It is suggested that habit plays a moderate role between intention and behavior. When habit was weak, intention had a stronger effect on behavior of choice in intercity modes, but when habit was strong there was relative weak relationship between intention and behavior, supporting the sixth hypothesis.

### 7. Conclusions

The paper introduces the theory of planned behavior into the research of intercity travel mode choice and includes descriptive norms and habit as new latent factors into the theory. The choice and intention of the three intercity travel modes in Yangtze River Delta can be explained by the original predictors of TPB. Moreover, descriptive norm and habit may increase significant explained variance in intention. Particularly, introduction of habit in hierarchical regression analyses results in the biggest incremental explained variance in intention. Habit not only has a significant effect on intention but also operates as moderator of intention on behavior of mode choice. The paper also constructs the MIMIC models to research the relationship between demographic characteristics and the latent factors of the expanded TPB. The results show that socioeconomic statuses of travelers have different significant impact on the latent factors. In addition, the analyses among the latent factors in the expanded TPB verify that the theory's suitability of intercity travel mode choice and increase the understanding of the role of descriptive norms and habit in TPB. Based

on the understanding and the different demographic statistical characteristics, transportation planners could design a socially desirable sustainable transportation system in line with people's preferences.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This research was supported by MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project no. 11YJC630084), Jiangsu Social Science Fund (Project no. 11SHD011), and Transportation Construction Project of Science and Technology (Project no. 2011318321960).

### References

- [1] C. R. Bhat, "A heteroscedastic extreme value model of intercity travel mode choice," *Transportation Research B*, vol. 29, no. 6, pp. 471–483, 1995.
- [2] J. Zhao, "Preference accommodating and preference shaping: incorporating traveler preferences into transportation planning," Massachusetts Institute of Technology, 2009.
- [3] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179–211, 1991.
- [4] I. Ajzen, "The theory of planned behaviour: reactions and reflections," *Psychology and Health*, vol. 26, no. 9, pp. 1113–1127, 2011.
- [5] I. Ajzen, "Constructing a TPB questionnaire: conceptual and methodological considerations," pp. 1–13, 2002, <http://people.umass.edu/ajzen/pdf/tpb.measurement.pdf>.
- [6] B. Gardner, "Modelling motivation and habit in stable travel mode contexts," *Transportation Research F*, vol. 12, no. 1, pp. 68–76, 2009.
- [7] A. Rivas and P. Sheeran, "Descriptive norms as an additional predictor in the theory of planned behaviour: a meta-analysis," *Current Psychology*, vol. 22, no. 3, pp. 218–233, 2003.
- [8] R. B. Cialdini, C. A. Kallgren, and R. R. Reno, "A focus theory of normative conduct: a theoretical refinement and reevaluation of

- the role of norms in human behavior," *Advances in Experimental Social Psychology*, vol. 24, pp. 201–234, 1991.
- [9] B. Gardner and C. Abraham, "Psychological correlates of car use: a meta-analysis," *Transportation Research F*, vol. 11, no. 4, pp. 300–311, 2008.
- [10] Y. Heath and R. Gifford, "Extending the theory of planned behavior: predicting the use of public transportation," *Journal of Applied Social Psychology*, vol. 32, no. 10, pp. 2154–2189, 2002.
- [11] J. Thøgersen, "Norms for environmentally responsible behaviour: an extended taxonomy," *Journal of Environmental Psychology*, vol. 26, no. 4, pp. 247–261, 2006.
- [12] K. S. Courneya, M. Conner, and R. E. Rhodes, "Effects of different measurement scales on the variability and predictive validity of the "two-component" model of the theory of planned behavior in the exercise domain," *Psychology and Health*, vol. 21, no. 5, pp. 557–570, 2006.
- [13] H. Aarts, B. Verplanken, and A. van Knippenberg, "Predicting behavior from actions in the past: repeated decision making or a matter of habit?" *Journal of Applied Social Psychology*, vol. 28, no. 15, pp. 1355–1374, 1998.
- [14] I. Ajzen, "Residual effects of past on later behavior: habituation and reasoned action perspectives," *Personality and Social Psychology Review*, vol. 6, no. 2, pp. 107–122, 2002.
- [15] S. Wu and T. Lin, "Exploring knowledge sharing behavior of IS personnel with theory of planned behavior," *Journal of Information Management*, vol. 14, no. 2, pp. 75–110, 2007.
- [16] T. Yu, N. Lee, and G. Wu, "An empirical research of Kinmen tourists' behavioral tendencies model—a case of cross-validation in causal modeling," *Journal of Tourism Studies*, vol. 11, no. 4, pp. 355–384, 2005.
- [17] C. Tsai, "Applying the theory of planned behavior to explore the independent travelers' behavior," *African Journal of Business Management*, vol. 4, no. 2, pp. 221–234, 2010.
- [18] D. Mahon, C. Cowan, and M. McCarthy, "The role of attitudes, subjective norm, perceived control and habit in the consumption of ready meals and takeaways in Great Britain," *Food Quality and Preference*, vol. 17, no. 6, pp. 474–481, 2006.
- [19] B. Verplanken, H. Aarts, A. van Knippenberg, and A. Moonen, "Habit versus planned behaviour: a field experiment," *The British Journal of Social Psychology*, vol. 37, no. 1, pp. 111–128, 1998.
- [20] J. C. Nunnally, I. H. Bernstein, and J. M. T. Berge, *Psychometric Theory*, vol. 2, McGraw-Hill, New York, NY, USA, 1967.
- [21] J. M. Cortina, "What is coefficient alpha? An examination of theory and applications," *Journal of Applied Psychology*, vol. 78, no. 1, pp. 98–104, 1993.
- [22] R. E. Rhodes, R. C. Plotnikoff, and J. C. Spence, "Creating parsimony at the expense of precision? Conceptual and applied issues of aggregating belief-based constructs in physical activity research," *Health Education Research*, vol. 19, no. 4, pp. 392–405, 2004.
- [23] B. Verplanken and S. Orbell, "Reflections on past behavior: a self-report index of habit strength," *Journal of Applied Social Psychology*, vol. 33, no. 6, pp. 1313–1330, 2003.
- [24] S. Bamberg, I. Ajzen, and P. Schmidt, "Choice of travel mode in the theory of planned behavior: the roles of past behavior, habit, and reasoned action," *Basic and Applied Social Psychology*, vol. 25, no. 3, pp. 175–187, 2003.

## Research Article

# Conspicuity Research on the Highway Roadside Objects: A Simulator Study

**Xuemei Chen,<sup>1</sup> Zhonghua Wei,<sup>2</sup> Xia Zhao,<sup>2</sup> Mingyang Hao,<sup>2</sup> and Tongyang Zhang<sup>2</sup>**

<sup>1</sup> School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

<sup>2</sup> College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Xia Zhao; 386931448@qq.com

Received 15 November 2013; Revised 20 January 2014; Accepted 21 January 2014; Published 5 March 2014

Academic Editor: Huimin Niu

Copyright © 2014 Xuemei Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In a monotonous travelling environment, the single-vehicle run-off-roadside accidents occur easily. The injuries and fatalities caused by those accidents are significant components of the annual road casualties. The causation is the complex interaction of the visual effects on the roadside objects' conspicuity. So the conspicuity enhancement needs to be considered in the roadside objects design to provide a temporary restoration of alertness and vigilance to drivers. Factors contributing to the conspicuity of the roadside objects were analyzed in this paper. A driving simulator study was conducted in order to extrapolate the relationship between the legibility distances and the objects and to quantify the conspicuity of the roadside objects different in basic features. The conclusions of this paper were firstly, a significant correlation existed between the mean legibility distance and the object's size. The mean legibility distance was in a significant exponential proportion to the object's size. Secondly, the triangle's legibility was better than that of the rectangle and round contours. Only when the roadside object was combined with the suitable contour and size did the best visual quality come. To some extent, the conclusions could provide theoretical tools and strategies to optimize the dimensional design of the roadside objects in order to maintain the roadside safety.

## 1. Introduction

A typical highway driving task may not contain a complex cognitive process, but in a monotonous travelling environment, the single-vehicle run-off-roadside accidents occur easily [1]. Most of the accidents are collisions to the roadside human-made objects, such as the ditches, utility poles, crash cushions, embankments, or signposts. The injuries and fatalities caused by those collisions are significant components of annual road casualties. For example, in the United States, the single-vehicle run-off-roadside accidents result in over a million highway crashes with the roadside objects every year. The latest accident statistics in Washington State indicated that roadside crashes account for one-third of the total highway fatalities and about one-fourth of the traffic accidents were associated with vehicles running off the road on the highways. Such accidents accounted for about one-third of all highway fatalities, with an estimated societal cost of over \$80 billion a year [2]. Also, the accident statistics in

the European Union in 1998 indicated that 33.8 percent of all fatalities occurred when the single vehicles left the roadside unintentionally and approximately two-thirds of fatalities on the rural roads were caused by it [3, 4].

The causation of these accidents is the complex interaction of the visual effects on the roadside objects' conspicuity. Among the run-off-roadside accidents, most were associated with the driving inattention and drowsiness, which triggered the visual distractions, visual fatigue, or looking-but-failing-to-see faults [5]. Loo [6] noted that the field dependent drivers were less skilled at detecting the roadside objects embedded in visual scenes, especially when the roadside objects were in the unobtrusive contours in a monotonous highway environment [7]. Several findings showed that the frequencies of the run-off-roadside accidents could be significantly reduced by enhancing the roadside objects' conspicuity [8]. So the roadside objects should be in great conspicuity to ensure the physical visual condition [9–11]. The conspicuity enhancement needs to be considered in the dimensional

design of the roadside artificial objects to provide a temporary restoration of alertness and vigilance to the drivers [12]. But within the international community, no clear agreement has been reached on how these roadside objects should be dimensioned [13]. This further highlights the task to quantify the conspicuity of the roadside objects [14, 15].

Studies to quantify the conspicuity can be linked to the studies on the objective legibility distance and the subjective comfort preference of the viewers. Factors contributing to the roadside objects' legibility are their basic features and the context in which they are embedded [16–21]. The basic features include the sizes, the colors, and the contours of the roadside objects. And the embedded context includes the contrast with the immediate surroundings, their placement, and the complex interaction with the traffic background and the drivers' requirements or expectations. Subjective comfort assessment on the conspicuity can be quantified via the questionnaires.

So the objective of this exploratory study was to quantify the conspicuity of the roadside objects. To do this, two primary steps must be done: firstly, to extrapolate the relationship between the legibility distances and the objects and, secondly, to assess the subjective comfort preference of the roadside objects varied in the basic features. The conclusions could provide theoretical tools and strategies to optimize the dimensional design of roadside objects. The rest of the paper was organized as follows. Section 2 introduced the method to conduct the conspicuity test. And Section 3 analyzed the experimental data of the test. And conclusions and discussions were given in Section 4.

## 2. Method

**2.1. Participants.** Forty participants were recruited to drive in the simulated scenario. 28 of them were males. The mean age was 26 years ( $SD = 2$ ). Each participant was a licensed driver with at least 3 years' driving experience. All of them had a minimum visual acuity of 20/40, wearing corrective lenses if necessary. Psychoactive substances, such as the caffeine or nicotine, were forbidden to be taken before or during the formal experiment. Each participant received the monetary compensation for their involvement in the experiment.

**2.2. Apparatus.** The chronic lack of detailed data on the conspicuity of the roadside objects has been an obstacle to conduct the conspicuity research on the roadside objects. Supposing the data on the conspicuity can be collected, the best option is to conduct this investigation in a real car travelling through a real highway section. However, such trials are not replicable, as the identical circumstances cannot be replicated for other tests because of the changing traffic and weather. So the controlled situations for the conspicuity test are preferred. A driving simulator, an interactive device based on PC programs, is recognized as an effective tool to conduct the conspicuity test [22, 23]. The external interfering elements can be eliminated or controlled just by the simple parameter modification in the scenarios [24–26]. Studies have proven that the legibility distance measured via a



FIGURE 1: A tester drove in the simulator.

simulator agrees with the field legibility distance [27]. In addition, the subjective comfort preference of the roadside objects can be easily measured to quantify the conspicuity [28].

The simulator AutoSim AS 1600 used in this study is a fixed-based driving simulator composed of a complete automobile, fully functional pedals and dashboard, and a large screen showing highway images projected by an RGB projector. It provides a realistic visual, sound, and vibration system. Simple desktop simulators, which are often implemented using a standard PC computer monitor for the visual display, are on the other end of the range. During a simulation test, the location and speed of the vehicle on the  $x$ -,  $y$ -, and  $z$ -axes are recorded by the GPS tracking software and displayed on the GPS real-time tracking device window. And a loudspeaker is mounted on the wheel to receive the driver's real-time feedback. Room temperature and lighting are controlled. Figure 1 showed that a tester drove in the simulator.

**2.3. Simulation Scenarios.** The simulation scenarios in this study were based on a 10 km-long corridor of the Jing-jin highway, which is started from the 4th east ring of Beijing, China. It is a six-lane state highway containing a side slope with typical grassland covered on the topography (as shown in Figure 2). The lane widths are obtained from the road controlling authority, and the road markings are consistent with Chinese Transport Agency guidelines. The driving speed is limited to 120 km/h.

Due to the high-traveling speed, all the roadside objects are identified by the spatial contours, which is the key information for drivers to process when driving. Three contours, the triangle, the rectangle, and the round, frequently appear on the roadside as the sign supports, bridges, cut-type slopes, ditches, fences, utility poles, guardrails, crash cushions, breakaway posts, or embankments and so forth [29]. Accordingly, three simulated scenarios, classified by these three roadside contours, were created. In each scenario, five roadside objects, which were in the identical contours



FIGURE 2: Color contrast on the conspicuity.

but varied sizes, were structured on the center of the slope in sequence. And the interval between each two adjacent objects under the limited speed was calculated according to the short-term memory time—the duration time for holding the relevant information active in mind for 60 seconds at most [30]. The design parameters of the three scenarios were described in Table 1.

To quantify the conspicuity of the roadside objects, the embedded context, including the objects' placement and their color contrast with the immediate surroundings, was supposed to be kept in a conspicuous level, so that the emphasis could be put on testing the conspicuity of the roadside objects without interference. Several hypotheses were made on those factors to satisfy the basic conspicuity in the scenarios.

- (1) Set the roadside objects in a sensitive visual zone. The most sensitive distance off the driveway is testified to be ranged from 15 to 25 meters [31]. In this paper, the distance off the driveway to the roadside objects was 15 meters.
- (2) Set the color contrast with the immediate surroundings conspicuously. Color contrast refers to the perceptual difference in two adjacent colors. Most studies support the notion that color contrast has only a small effect on the conspicuity [32–34]. The 24-color wheel theory (as shown in Figure 3) identifies that conspicuity appears when the colors exactly opposite to each other are combined together [35]. So, in this paper, the objects were textured in red purple in contrast to the green slope to bring a maximum effect on the conspicuity.

**2.4. Procedure.** During the preexperiment, each tester was given a brief overview of the experimental activities and practiced for 10–15 minutes to get familiar with the scenarios and rules. At the formal experiment of each scenario, the tester was instructed to drive in the simulator. Factors contributing to the conspicuity—the dynamical legibility distance, in addition to the comfort assessment on the roadside object—were measured under four given speeds, 60 km/h, 80 km/h, 100 km/h, and 120 km/h. A short break was offered after each scenario in case of the driving fatigue.

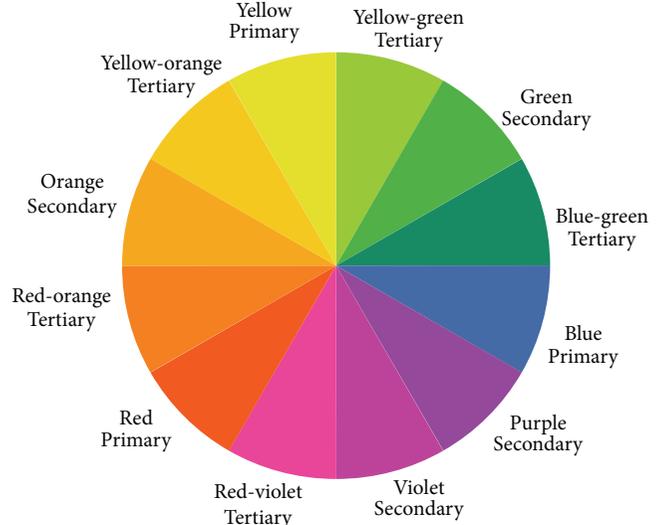


FIGURE 3: The 24 color wheel.

**2.5. Dependent Variables.** The in-vehicle data collection system provided the capability to store data on a computer in the form of one line of numerical data every 0.1 second during a data run. The specific measures collected were as follows.

(1) *The Legibility Distance.* In order to test the dynamical legibility distance, the tester was asked to drive towards the object, focus on the roadside object, and try to identify its features. Once having identified it, he/she was instructed to sound the horn and named it loudly. Then one experimenter recorded the point of the vehicle, named  $S_1$ , by its axes. The next point of the vehicle, named  $S_2$ , was recorded when the tester passed the object. The legibility distance, named  $S$ , was calculated in the following way:

$$S = |S_2 - S_1|. \tag{1}$$

(2) *The Comfort Assessment on the Conspicuity.* The tester was asked via the loudspeaker about their comfort assessment of each object during his/her driving process, which was quantified via a Likert scale ranged from 1 “not comfortable” to 7 “extremely comfortable.” Considering the fact that the assessment was given under a dynamical driving process, each tester needed to get familiar with it during the pre-experiment. And the answer was recorded correctly on the questionnaire sheet. The total assessment costs 30 seconds at most.

### 3. Results

The data on the legibility distances and the comfort assessments under given speed and contours were used to analyze the conspicuity of the roadside objects in detail.

**3.1. Legibility Distance.** The data on the mean legibility distances of the 15 roadside objects were depicted in Tables 2, 3, 4, and 5. Firstly, the Pearson correlation test was

TABLE 1: The design parameters of the roadside objects.

Simulated scenario	Contours	Sizes (m <sup>2</sup> )					Interval (km)	Highway length (km)
1	Rectangle	1	16	49	100	169	2	10
2	Triangle	1	16	49	100	169	2	10
3	Round	1	16	49	100	169	2	10

TABLE 2: The mean legibility distance under 60 km/h.

Contour	Sizes of the roadside objects (m <sup>2</sup> )				
	1	16	49	100	169
Triangle	136.58	529.10	846.98	1225.88	1761.12
Rectangle	123.93	514.48	762.77	1129.70	1606.03
Round	103.63	493.01	647.66	1128.78	1559.14

conducted to check the correlation between the mean legibility distance, the size, and the driving speed. The results presented in Table 6 showed that no significant correlation existed between the mean legibility distance and the speed; but a significant correlation existed between the mean legibility distance and the size of the contours.

The mathematical relationship between the mean legibility distance and the size under different speed was analyzed by SPSS 16.0. Functions between them were given in Figures 4, 5, 6, and 7. Every function curve was monotone increasing, with a high fitting of  $R^2$ . Moreover, the mean legibility distance was in a significant exponential proportion to the size.

Under speed 120 km/h, the legibility distance of each object was converted into the legibility assessment via a Likert scale ranged from 1 “not legible” to 7 “extremely legible”. Supposing the maximum legibility distance was evaluated to be 7 “extremely legible”, then other legibility assessments were normalized in the following way:

$$LA_{ij} = L_{ij} \times \frac{L_{\max}}{7}, \quad (2)$$

where  $LA_{ij}$  is the legibility assessment score on the object in  $i_{th}$  contour and  $j_{th}$  size;  $L_{ij}$  is the mean legibility distance of the object in  $i_{th}$  contour and  $j_{th}$  size;  $L_{\max}$  is the maximum value of the mean legibility distances among all the objects;  $i$  is the contour of the objects,  $i = 1, 2, 3$ ; and  $j$  is the sizes of the objects;  $j = 1, 2, \dots, 5$ .

So the legibility assessment results of all the roadside objects were presented in Table 7 and Figure 8. It is depicted that the triangle contour in size 169 m<sup>2</sup> won the highest legibility score, while the round contour in size 1 m<sup>2</sup> was rated the least legible with an average rating of 0.42. The same results were found under the speeds 60 km/h, 80 km/h, and 100 km/h. The triangle’s legibility was testified to be the best among the three given contours in same sizes.

**3.2. Comfort Assessment.** Under the speed 120 km/h, the comfort assessment regarding the experimental roadside object was fulfilled via the in-driving questionnaire, with the results shown in Table 8 and Figure 9. It is seen that the rectangle contour in size 100 m<sup>2</sup> won the best comfort

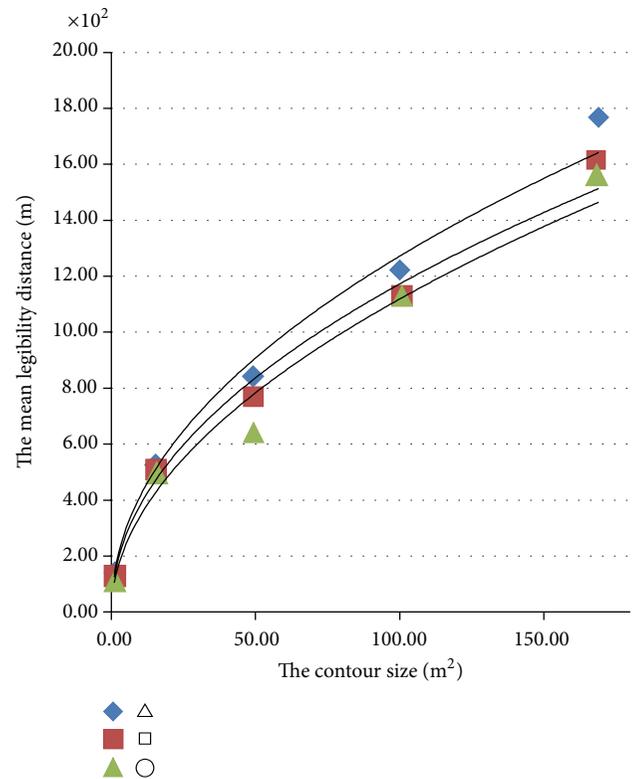


FIGURE 4: Relationship between the mean legibility distance and the contour size under 60 km/h, where the relationship between the mean legibility distance and the contour size under 60 km/h for triangle:  $y = 135x^{0.486}$  and  $R^2 = 0.907$ , and for round:  $y = 105x^{0.513}$  and  $R^2 = 0.908$ .  $x$  represents the size of the roadside object, unit is square meters,  $y$  represents the mean legibility distance, unit is meter, and  $R^2$  represents the degree of fitting, describing how well it fits a function.

preference among all the roadside objects with an average of 6.1, while the triangle contour in size 1 m<sup>2</sup> won the poorest comfort preference with an average rating of 0.36. The same results were found under the speeds 60 km/h, 80 km/h, and 100 km/h. When the sizes were smaller than 100 m<sup>2</sup>, the rectangle and round contours could easily win the

TABLE 3: The mean legibility distance under 80 km/h.

Contours	Sizes of the roadside objects (m <sup>2</sup> )				
	1	16	49	100	169
Triangle	122.87	524.56	766.5	1397.16	1764.90
Rectangle	117.65	516.37	729.91	1236.19	1561.38
Round	107.87	505.81	696.44	1202.48	1517.41

TABLE 4: The mean legibility distance under 100 km/h.

Contours	Sizes of the roadside objects (m <sup>2</sup> )				
	1	16	49	100	169
Triangle	123.72	532.65	757.29	1347.34	1764.90
Rectangle	120.82	524.62	744.64	1248.22	1744.74
Round	113.86	500.81	732.43	1162.23	1642.60

TABLE 5: The mean legibility distance under 120 km/h.

Contours	Sizes of the roadside objects (m <sup>2</sup> )				
	1	16	49	100	169
Triangle	121.38	514.18	756.34	1331.08	1776.47
Rectangle	112.77	507.39	749.48	1212.79	1753.18
Round	106.43	482.58	717.82	1065.30	1701.53

TABLE 6: Pearson correlation analysis of the variables.

	Legibility distance	Contour size	Driving speed
Legibility distance			
Pearson correlation	1.000	0.977	0.018
Significance	—	0.000	0.893
Contour size			
Pearson correlation	0.977	1.000	0.000
Significance	0.000	—	1.000
Driving speed			
Pearson correlation	0.018	0.000	1.000
Significance	0.893	1.000	—

TABLE 7: The legibility assessment scores of the objects under 120 km/h.

Contours	Sizes of the roadside objects (m <sup>2</sup> )				
	1	16	49	100	169
Triangle	0.48	2.03	2.98	5.24	<b>7.00</b>
Rectangle	0.44	2.00	2.95	4.78	6.91
Round	0.42	1.90	2.83	4.20	6.70

\* 1: not at all legible. 7: extremely legible.

TABLE 8: The comfort assessment scores of the roadside objects under 120 km/h.

Contours	Sizes of the roadside objects (m <sup>2</sup> )				
	1	16	49	100	169
Triangle	0.36	2.46	4.25	5.8	2.48
Rectangle	0.40	3.07	4.68	5.9	3.02
Round	0.45	3.58	5.02	6.0	3.28

\* 1: not at all comfortable. 7: extremely comfortable.

TABLE 9: The conspicuity assessment scores of the objects under 120 km/h.

Contours	Sizes of the roadside objects (m <sup>2</sup> )				
	1	16	49	100	169
Triangle	0.43	2.20	3.49	5.47	5.19
Rectangle	0.43	2.43	3.64	5.23	5.35
Round	0.43	2.57	3.71	4.92	5.33

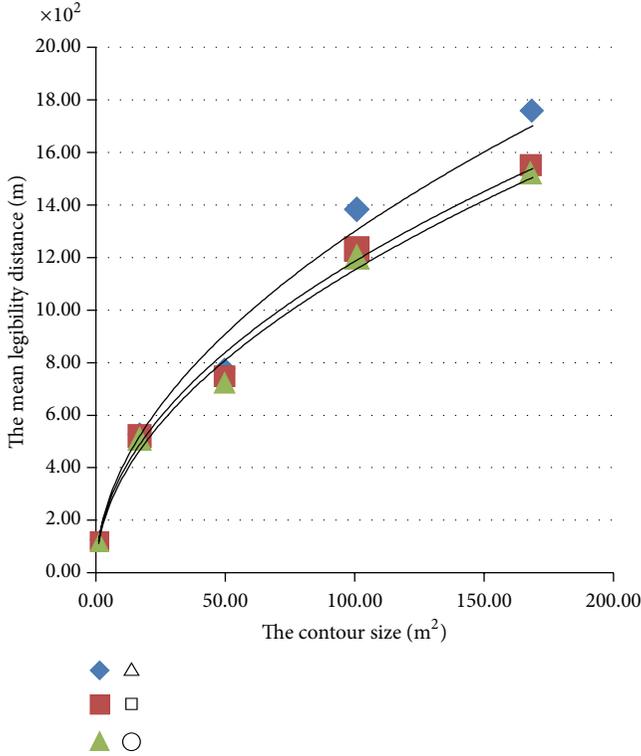


FIGURE 5: Relationship between the mean legibility distance and the contour size under 80 km/h, where the relationship between the mean legibility distance and the contour size under 80 km/h was for triangle:  $y = 121.1x^{0.515}$  and  $R^2 = 0.892$ , and for round:  $y = 110.2x^{0.509}$  and  $R^2 = 0.902$ .

comfort preference, not the triangle contour, but the comfort preference was not monotone increasing to the sizes. When the sizes of the objects were larger than 100 m<sup>2</sup>, the comfort preference decreased gradually.

**3.3. Conspicuity Result.** The overall conspicuity was up to both the objective legibility distance and the subjective comfort preference. The overall conspicuity was calculated from the sum of the weighted normalized score assessment representing the factors in the following way:

$$C_{ij} = \alpha \times (LA_{ij}) + (1 - \alpha) \times (CA_{ij}), \quad (3)$$

where  $C_{ij}$  is the conspicuity score on the object in  $i_{th}$  contour and  $j_{th}$  size;  $LA_{ij}$  is the legibility assessment score on the object in  $i_{th}$  contour and  $j_{th}$  size;  $CA_{ij}$  is the comfort assessment score on the object in  $i_{th}$  contour and  $j_{th}$  size; and

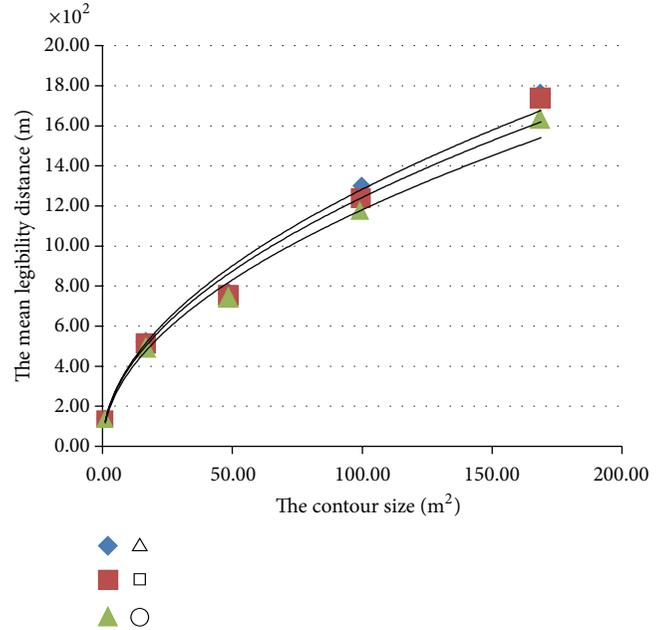


FIGURE 6: Relationship between the mean legibility distance and the contour size under 100 km/h, where the relationship between the mean legibility distance and the contour size under 100 km/h was for triangle:  $y = 122.5x^{0.51}$  and  $R^2 = 0.902$ , for rectangle:  $y = 120.2x^{0.507}$  and  $R^2 = 0.903$ , and for round:  $y = 114.2x^{0.507}$  and  $R^2 = 0.904$ .

$\alpha$  is the weight assigned to the legibility distance variable; in this paper  $\alpha = 0.6$ .

According to (3), the conspicuity results of the roadside objects based on the legibility and the comfort assessment were shown in Table 9 and Figure 10. It is seen that the triangle contour in size 100 m<sup>2</sup> was found to be the most conspicuous object with the comfort preference. Although the triangle contour, compared with the rectangle and round contours, was more conspicuous, its comprehensive conspicuity score was reduced because of its uncomfortable appearance. The same results were found under the speeds 60 km/h, 80 km/h, and 100 km/h.

## 4. Conclusions and Discussions

Several conclusions were given in this section to summarize the investigation of this paper. Firstly, the results showed that no significant correlation existed between the mean legibility distance and the speed, but a significant correlation existed between the mean legibility distance and the object's size.

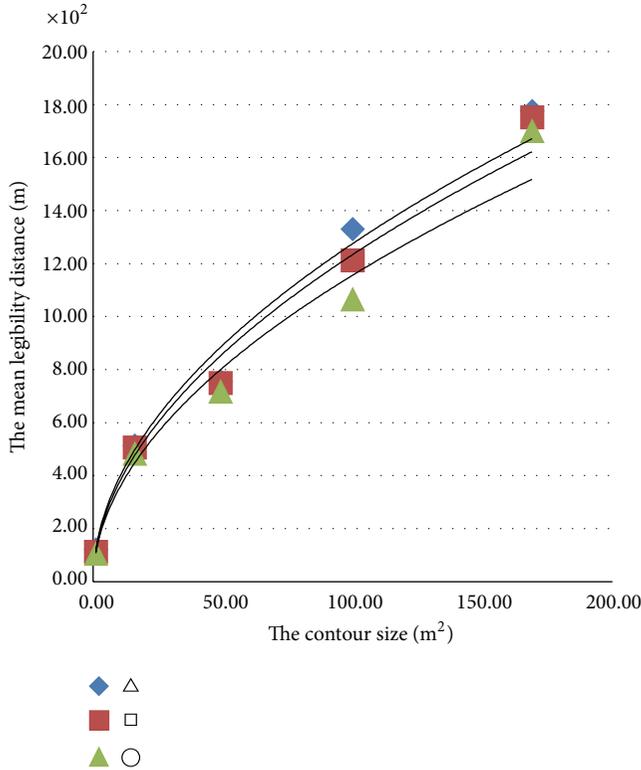


FIGURE 7: Relationship between the mean legibility distance and the contour size under 120 km/h, where the relationship between the mean legibility distance and the contour size under 120 km/h was for triangle:  $y = 119.2x^{0.514}$  and  $R^2 = 0.902$ , for rectangle:  $y = 112.4x^{0.520}$  and  $R^2 = 0.904$ , and for round:  $y = 106.4x^{0.518}$  and  $R^2 = 0.921$ .

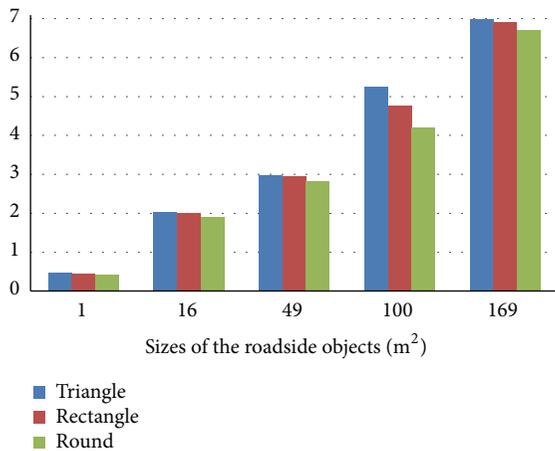


FIGURE 8: The legibility assessment of the objects under 120 km/h.

The mean legibility distance was in a significant exponential proportion to the object's size. Secondly, the comprehensive conspicuity of the roadside objects different in contours and sizes was quantified by the objective or subjective measurement on the objects' legibility and the comfort preference. The results confirmed the empirical theory that the triangle's legibility but not the conspicuity was better than those of

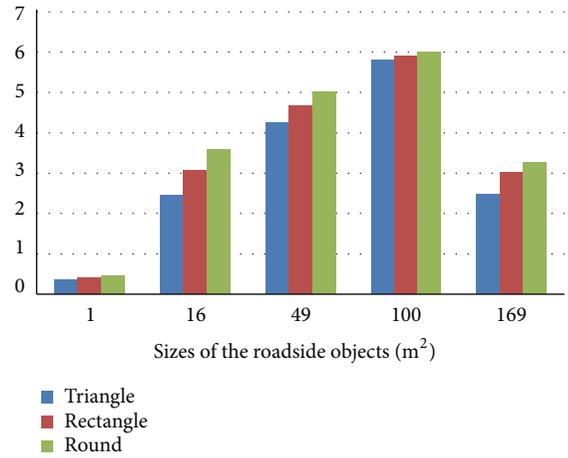


FIGURE 9: The comfort assessment of the roadside objects under 120 km/h.

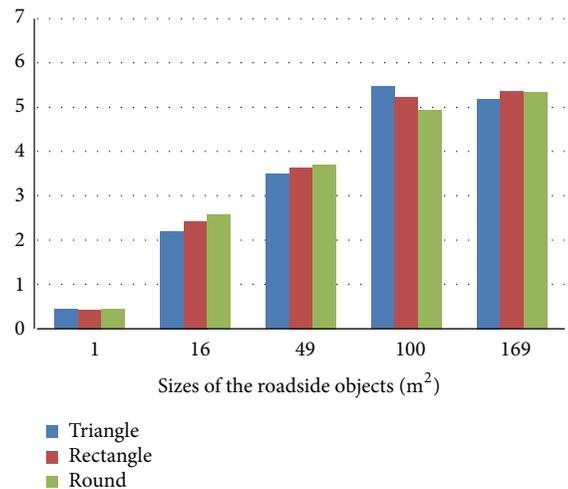


FIGURE 10: The conspicuity assessment of the objects under 120 km/h.

the rectangle and round contours. Only when the roadside object was combined with the suitable contour and size did the best visual quality come. The rectangle or round contours were preferred in the design and management of the roadside objects when the object was in a huge size. To some extent, the conclusions could provide theoretical tools and strategies to optimize the dimensional design of the roadside objects in order to maintain the roadside safety.

Further research is to be conducted to enrich its research scope.

Firstly, the experiment could be conducted among the elders because the current mean age of the participants was 26 years old. The effects of age on information-processing time and short-term memory stressed the elders due to a reduction in the minimum visibility distance and perception-reaction time required for certain signing situations. So the research plans should also be made to test the effects on older drivers of the roadside objects.

Secondly, besides the several basic features of the roadside objects (such as contours and sizes), other factors contributing to the conspicuity should also be considered in the simulated scenario. The color of the roadside object played an important role in conveying conspicuity. Therefore, it would be more useful to consider the effect of colors on the traffic safety. Additionally, the distinct contours of roadside structures should have a positive effect on the driving performance by allowing a temporary restoration of alertness and vigilance or leading to increased awareness of traffic route information, or enhancing the driving comfort due to the earlier detection of roadside objects and better visibility, although the novelty effects due to contours might wear off over time. In this paper, only three simple common contours of the roadside objects were selected; other complicated and monogram objects in contours could be investigated on the conspicuity further.

Thirdly, it is important to note that the real world varies from the simulator. There are gaps between the simulated and actual traffic scenarios, such as the external traffic disturbance, surroundings, and road condition. So the validation and modification work on several parameters will be conducted in actual traffic scenarios further.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgment

This work is supported by National Natural Science Foundation of China (Grant no. 51208008). Their support is gratefully acknowledged.

### References

- [1] C. Castro, *Human Factors of Visual and Cognitive Performance in Driving*, CRC Press, New York, NY, USA, 2008.
- [2] NCHRP, *Strategies for Improving Roadside Safety*, National Research Council Transportation Research Board, Washington, DC, USA, 1997.
- [3] C. Collin, Statistics in Focus-transport. Eurostat catalogue number CA-NZ-00-003-EN-I, 2000.
- [4] K. Johnson, "How to prevent run-off-the-road crashes," *Traffic Safety*, vol. 3, pp. 17–19, 2000.
- [5] S. D. Baulk, S. N. Biggs, K. J. Reid, C. J. van den Heuvel, and D. Dawson, "Chasing the silver bullet: Measuring driver fatigue using simple and complex tasks," *Accident Analysis and Prevention*, vol. 40, no. 1, pp. 396–402, 2008.
- [6] R. Loo, "Individual differences and the perception of traffic signs," *Human Factors*, vol. 20, no. 1, pp. 65–74, 1978.
- [7] E. Ayati, P. N. Mohammad, and A. S. Ali, "Introducing roadside hazard severity indicator based on evidential reasoning approach," *Safety Science*, vol. 50, pp. 1618–1626, 2012.
- [8] J. Lee and F. Mannering, "Analysis of roadside accident frequency and severity and roadside safety management," *Accident Analysis and Prevention*, vol. 34, no. 2, pp. 149–161, 2002.
- [9] K. Mak, "Safety effects of roadway design decisions—roadside," *Transportation Research Record*, no. 1512, pp. 16–21, 1995.
- [10] M. C. Good, J. C. Fox, and P. N. Joubert, "An in-depth study of accidents involving collisions with utility poles," *Accident Analysis and Prevention*, vol. 19, no. 5, pp. 397–413, 1987.
- [11] D. Wolford and D. L. Sicking, "Guardrail need: embankments and culverts," *Transportation Research Record*, no. 1599, pp. 48–56, 1997.
- [12] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 1, pp. 32–64, 1995.
- [13] J. Lee and F. Mannering, "Impact of roadside features on the frequency and severity of run-off-roadway accidents," *Accident Analysis and Prevention*, vol. 34, no. 2, pp. 149–161, 2002.
- [14] B. L. Cole and P. K. Hughes, "A field trial of attention and search conspicuity," *Human Factors*, vol. 26, no. 3, pp. 299–313, 1984.
- [15] S. E. Jenkins and B. L. Cole, "Daytime conspicuity of road traffic control devices," *Transportation Research Record*, pp. 74–80, 1986.
- [16] A. de la Escalera, J. M. Armingol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," *Image and Vision Computing*, vol. 21, no. 3, pp. 247–258, 2003.
- [17] D. M. Burns and N. L. Johnson, "Characterizing the visual performance of fluorescent retro-reflective signing materials using the fluorescent luminance factor, YF," in *Proceedings of the 8th Congress of the International Color Association*, vol. 1, pp. 359–362, 1997.
- [18] H. T. Zwahlen, Schnell, and T. Visual, *Detection and Recognition of Fluorescent Color Targets versus Non-Fluorescent Color Targets as a Function of Peripheral Viewing Angle and target Size*, National Research Council Transportation Research Board, Washington, DC, USA, 1997.
- [19] P. Thiffault and J. Bergeron, "Monotony of road environment and driver fatigue: a simulator study," *Accident Analysis and Prevention*, vol. 35, no. 3, pp. 381–391, 2003.
- [20] R. L. Anders, *On-road investigation of fluorescent sign colors to improve conspicuity [Ph.D. thesis]*, Virginia Polytechnic Institute and State University, Blacksburg, Va, USA, 2000.
- [21] G. Cook, G. Webber, S. A. Gillham, S. A. le Scouiller, and E. Moseley, "The legibility and conspicuity of emergency escape route signage for people with visual impairments," *International Congress Series*, vol. 1282, pp. 1016–1020, 2005.
- [22] S. de Groot, J. C. F. de Winter, J. M. L. García, M. Mulder, and P. A. Wieringa, "The effect of concurrent bandwidth feedback on learning the lane-keeping task in a driving simulator," *Human Factors*, vol. 53, no. 1, pp. 50–62, 2011.
- [23] J. C. F. de Winter, S. de Groot, M. Mulder, P. A. Wieringa, J. Dankelman, and J. A. Mulder, "Relationships between driving simulator performance and driving test results," *Ergonomics*, vol. 52, no. 2, pp. 137–153, 2009.
- [24] W. Zhonghua, C. Xuemei, and G. Ming, "A study on guide sign validity in driving simulator," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1290–1297, 2011.
- [25] S. D. Baulk, S. N. Biggs, K. J. Reid, C. J. van den Heuvel, and D. Dawson, "Chasing the silver bullet: measuring driver fatigue using simple and complex tasks," *Accident Analysis and Prevention*, vol. 40, no. 1, pp. 396–402, 2008.
- [26] J. A. Telner, D. L. Wiesenthal, and E. Bialystok, "Video gamer advantages in a cellular telephone and driving task," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 53, no. 23, pp. 1748–1752, 2009.
- [27] A. Toet, F. L. Kooi, P. Bijl, and J. M. Valetton, "Visual conspicuity determines human target acquisition performance," *Optical Engineering*, vol. 37, no. 7, pp. 1969–1975, 1998.

- [28] R. W. Allen, G. D. Park, M. L. Cook, and D. Fiorentino, "The effect of driving simulator fidelity on training effectiveness," in *Proceedings of the Driving simulation conference*, Iowa City, Iowa, USA, 2007.
- [29] Research Institute of Highway Ministry of Transport, *Road Traffic Signs and Markings*, China Standards Press, Beijing, China, 2009.
- [30] "Short-term memory," [http://en.wikipedia.org/wiki/Short-term\\_memory](http://en.wikipedia.org/wiki/Short-term_memory).
- [31] Z. H. Wei, *Research on highway landscape design theory [Ph.D. thesis]*, Beijing University of Technology, Beijing, China, 2005.
- [32] D. Funkhouser, S. Chrysler, A. Nelson, and E. S. Park, "Traffic sign legibility for different sign background colors: results of an open road study at freeway speeds," *Proceedings of the Human Factors and Ergonomics Society. Annual Meeting*, vol. 52, no. 23, pp. 1855–1859, 2008.
- [33] P. L. Olson and A. Bernstein, "The nighttime legibility of highway signs as a function of their luminance characteristics," *Human Factors*, vol. 21, no. 2, pp. 145–160, 1979.
- [34] T. W. Forbes, J. P. Fry Jr., R. P. Joyce et al., *Letter and Sign Contrast, Brightness, and Size Effects on Visibility*, National Research Council Transportation Research Board, Washington, DC, USA, 1968.
- [35] "Introduction to color science," <http://www.itp.uni-hannover.de/~zawischa/ITP/introcol.html>.

## Research Article

# A Quasi-Poisson Approach on Modeling Accident Hazard Index for Urban Road Segments

**Lu Ma, Xuedong Yan, and Wenxin Qiao**

*MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Xuedong Yan; [xdyan@bjtu.edu.cn](mailto:xdyan@bjtu.edu.cn)

Received 31 October 2013; Revised 3 January 2014; Accepted 5 January 2014; Published 4 March 2014

Academic Editor: Huimin Niu

Copyright © 2014 Lu Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In light of the recently emphasized studies on risk evaluation of crashes, accident counts under specific transportation facilities are adopted to reflect the chance of crash occurrence. The current study introduces more comprehensive measure with the supplement information of accidental harmfulness into the expression of accident risks which are also named Accident Hazard Index (AHI) in the following context. Before the statistical analysis, datasets from various sources are integrated under a GIS platform and the corresponding procedures are presented as an illustrated example for similar analysis. Then, a quasi-Poisson regression model is suggested for analyses and the results show that the model is appropriate for dealing with overdispersed count data and several key explanatory variables were found to have significant impact on the estimation of AHI. In addition, the effect of weight on different severity levels of accidents is examined and the selection of the weight is also discussed.

## 1. Introduction

Aggregated accident analyses for certain transportation facilities including road segments, intersections, and recently emphasized traffic analysis zones (TAZs) have been thoroughly studied in the forms of accident frequency and accident rates [1, 2]. The major approach in those studies is to use regression models to make connections between number of accidents and attributes of these facilities, such as geometric and traffic characteristics of road segments and socioeconomic and demographic properties of TAZs. Past studies can readily consider many aspects of accident data as well as deal with some critical issues. For example, the earlier models based on the Poisson distribution actually rely on the nature of the Poisson process of accident occurrence [3], the negative binomial distribution improves the Poisson model by introducing the consideration of overdispersed data [4] and later more advanced zero-inflated models involve the specification of excessive zero counts [5] into the model. It is no doubt that these advanced models have discovered the nature of accident counts in different respects. However, for the purposes of traffic safety evaluation, the value of accident frequency or accident rate is less informative to

reflect the magnitude of accident risks because it cannot take the severity attributes of each accident into account.

Generally, a comprehensive risk function for safety evaluation should be able to measure the expected harmfulness of each accident and some practical analysis [6] raises such consideration for risk analyses. Such risk function actually contains two sources of information, namely, the chance of accident occurrence and harmfulness of each accident. Nevertheless, most previous studies neglect the severity information in measuring safety risks, and the exception would be several multivariate analyses in which a predefined distributional form of accident frequency for different levels of severity is adopted [7–9]. These multivariate models provide inference on the correlative relationships among accident frequencies for different severity levels, but they also cannot provide a single risk measurement for the purpose of safety evaluations. Towards this end, one of the objectives of this study seeks to define a single risk measure that can comprehensively capture the compound effect of accident occurrence and harmfulness on AHI.

Under this circumstance, this study also seeks to find appropriate regression models for predication and inference of AHI. With the consideration of possible overdispersion

on data, a quasi-likelihood model [10–12] is adopted as it provides a semiparametric method to estimate the mean value of interested parameters and hence it is a more nature approach for risk predications because less distributional assumptions are required. The suggested model can provide an important alternative to the frequently used negative binomial model. In addition, the quasi-Poisson model has been found to be more accurate for certain count data than the negative binomial models [13]. Besides, it is also necessary to present the procedure of data integration of this study as an example for similar analyses in the future. In fact, such analysis requires multiple data inputs [14] including accident data, traffic system data, road segments data, and traffic flow data from various sources.

In sum, one of the major objectives of this study is to identify a comprehensive measure for safety evaluation in terms of accident risks, namely, the Accident Hazard Index (AHI) in the following context. With such consideration, this study also tries to contribute a statistical analysis through the quasi-Poisson likelihood model which is suggested as a nature way to deal with overdispersed count data. The results will be analyzed by variables with significant coefficients and the effects of these variables on AHI are also illustrated. This study provides an alternative method to analyze the accident risk, in which the weight crash rate for different level of severity is used instead of the traditional analysis on the single value of crash rate. In addition, the selection of the value of weight is also presented as an indication to similar analyses.

## 2. Methods

**2.1. Datasets.** This research will focus on predication and inference on the Accident Hazard Index for road segments and the used accident data was collected in Pikes Peak Area, Colorado, USA, during the period from July 2006 to December 2010. In order to aggregate these accidents for road segments in this area, as well as to incorporate several key variables for regression, another dataset of road segments is also used in this study. The data describe each accident in terms of its severity, time of accident, locations, and so forth. The supplemental datasets describe the road traffic and geometric characteristics including variables such as the length of road segments, average annual daily traffic (AADT), ownership of the road segments, and number of through lanes. The traffic and roadway data is obtained from Colorado Department of Transportation (CDOT). The accident dataset is obtained from the Department of Revenue (DOR) and coded into GIS database by the Pikes Peak Area Council of Government (PPACG). With the location information of each accident, GIS platform can be used as a tool to integrate the two sources of data [15].

For demonstrative purposes, Figure 1 presents a sample area, which contains the road network map and the corresponding accidents. In this area, a road segment is highlighted and it can be used to illustrate the procedure for data integration. Through GIS platforms, the important step is to map accidents to the road segments to which they belong. Because accidents majorly occurred at either road

segments or intersections, the first step is to remove those accidents that occurred at intersections and a convenient approach considers an area with radius of 200 ft from the center of each intersection. Then for the remaining accidents, a 150 ft buffer size of double sides for road segments is used to consider the actual range of road segment and the observation error of location of accidents during the data collections. In addition, accidents occurred in ramps and parking lots are also excluded during the process.

Then with the constructed one-to-one corresponding mapping relationship between accidents and road segments, accidents for each road segment are aggregated by their severity levels. In this study, two levels of severity, namely, fatal-injury accidents and property-damage-only accidents, are considered. The distributions of accident counts by each level of severity are presented in Figure 2. Moreover several explanatory variables such as intersection density of road segments, urban/rural location, and ownership are considered in the following analysis. Table 1 provides the basic descriptive statistics for explanatory variables and the exposure variable.

**2.2. The Definition of Accident Hazard Index (AHI).** In order to measure accident risks with compound information of the likelihood of accident occurrence and corresponding harmfulness of each accident, a more general form for the Accident Hazard Index is presented as follows:

$$\begin{aligned} \Lambda &= E \left[ \frac{wN^I + (1-w)N^{II}}{t} \right] \\ &= wE \left( \frac{N^I}{t} \right) + (1-w)E \left( \frac{N^{II}}{t} \right), \end{aligned} \quad (1)$$

where  $\Lambda$  is the AHI that is expressed by the expected value of the weighted crash rates in terms of two severity levels. The measure is believed to have the ability to reflect the magnitude of accident risks for the purposes of safety evaluations as well as black spot diagnostics for road segments.  $N^I$  is the total number of fatal-injury accidents and  $N^{II}$  is the total number of property-damage-only accidents.  $w$  is the weight associated with fatal-injury accidents, which is ranging from 0.5 to 1 as a reflection of the relative importance of fatal-injury accidents in the analysis, and  $t$  is the exposure variable usually the vehicle mile traveled (VMT). Further, if  $w = 0.5$ , all types of accidents are treated equally and hence  $\Lambda$  will be proportional to crash rate and if  $w = 1$ , the property-damage-only accidents will be ignored.

In fact, it is true that fatal-injury accidents are usually accompanied with property damage. However, the crucial point to distinguish between the two types of accidents is that the property-damage-only accidents are defined to be accidents that are not fatal or injury-involved. Under such situation, even though fatal-injury accidents are usually associated with property damage, it can be assigned larger weight than property-damage-only accidents because it is reasonable to assume that the situation of fatal or injury-involved accident is more harmful than the situation of property-damage-only accident. As a result, the weighting

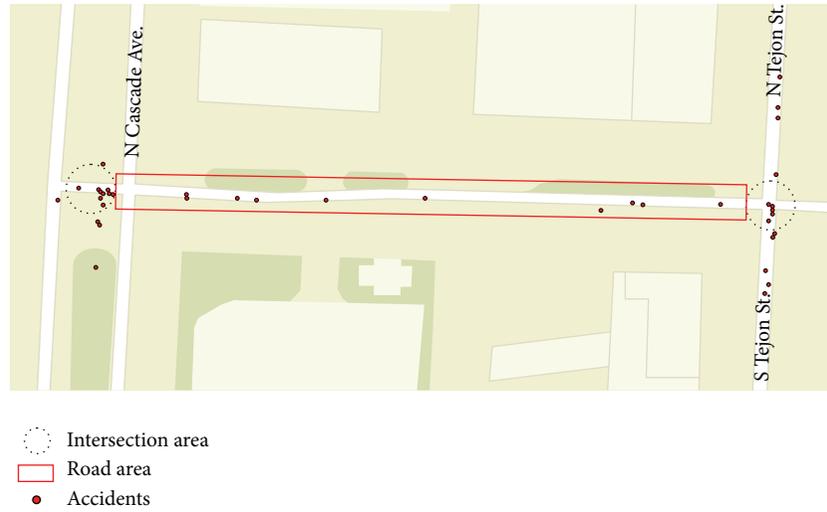


FIGURE 1: Illustration of the process of data integration.

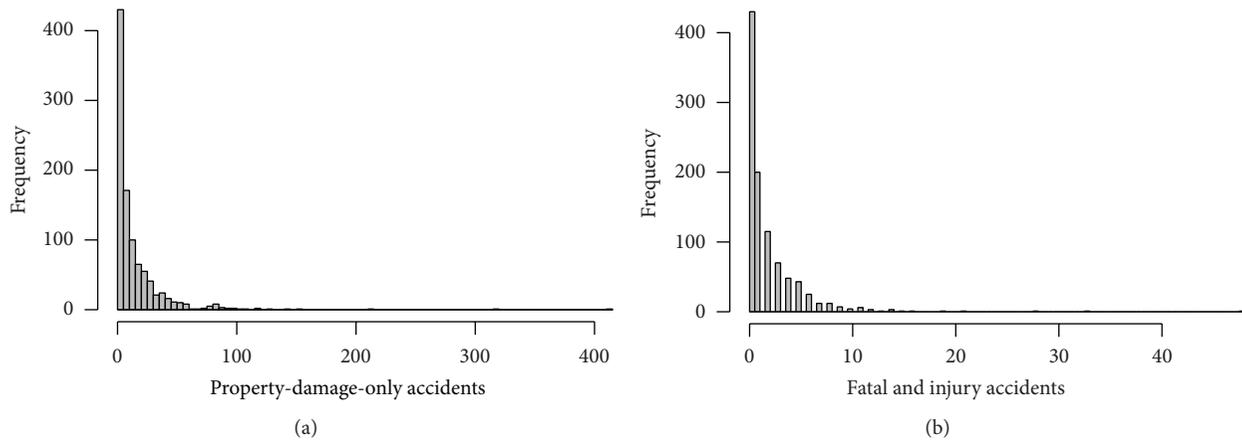


FIGURE 2: Histogram of accident counts.

TABLE 1: Descriptive characteristic for key variables.

Variables	Mean	Std. Dev.	Min.	Max.
The exposure variable (Offset)				
Vehicle miles traveled	11.760	48.236	0.037	797.400
Explanatory variables				
Intersection density	2.321	1.878	0	12.255
Road segment in rural areas	189		0	1
Road segment in urban areas	794		0	1
Ownership at state or federal levels	139		0	1
Ownership at other levels	844		0	1
Number of through lanes less and equal to 2	661		0	1
Number of through lanes greater than 2	322		0	1
Not on a designated truck route	893		0	1
Designated truck route	90		0	1
Annual average daily traffic	7523		40	53700
Pavement condition rating	2.997	0.737	1	5
Pavement type at low level	442		0	1
Pavement type at high level	541		0	1
Percent average daily trucks	2.400	2.477	1	17

process is meaningful to reflect the relative harmfulness of the two different types of accidents and therefore provide AHI as a more comprehensive and precise criterion indicating the overall losses from all types of accidents during the safety evaluations.

**2.3. The Quasi-Poisson Approach.** Accident Hazard Index in fact is the weighted expected value of two crash rates as defined above. It can be determined through the regression of  $N^I/t$  and  $N^{II}/t$ . As a result, the current study suggests quasi-likelihood models as an alternative to the traditional Poisson and negative binomial regression models since quasi-likelihood framework does not require a predefined distributional form of the responses and hence may produce more nature and accurate results [13]. In quasi-Poisson model, the variance is assumed to be the mean multiplied by a dispersion parameter. Therefore, the quasi-Poisson model is capable of considering overdispersed data, which is a common characteristic in accident counts. For assumed i.i.d. accident frequency  $N_1, N_2, \dots, N_m$  on road segments,  $\mu_i$  is the corresponding mean value such that

$$\mu_i = E(N_i). \quad (2)$$

In the following analysis, a log link function will be used. If  $t_i$  is denoted as the accident exposure measure for road segment  $i$ , the logarithm of  $t_i$  will be an offset term with fixed coefficient of one under the log link function as follows:

$$g(\mu_i) = \log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \log(t_i). \quad (3)$$

Instead of assumptions on the distributions, quasi-likelihood models only require specification of the relationship between mean and variance [16]. And the quasi-Poisson model adopts the relationship from the Poisson distribution such that the variance is related to mean only by a multiplication of the dispersion parameter  $\phi$  as follows:

$$\text{Var}(N_i) = \phi V(\mu_i) = \phi \mu_i, \quad (4)$$

where  $N_i$  is the number of accidents for certain type of severity and  $V(\mu_i)$  is the so-called variance function in a generalized linear model (GLM) setting. The quasi-score function  $q(\mu_i, n_i, \phi)$  is the first-order derivative of the log-likelihood function which is the same definition from the traditional score function. For the quasi-Poisson model, its score function for a single observation  $i$  is

$$q(\mu_i, n_i, \phi) = \frac{n_i - \mu_i}{\phi V(\mu_i)} = \frac{n_i - \mu_i}{\phi \mu_i}, \quad (5)$$

where  $n_i$  is the sample value for number of accidents in road segment  $i$ . Therefore, the quasi-likelihood function  $Q(\mu_i, n_i, \phi)$  for sample  $i$  can be written in the following form:

$$Q(\mu_i, n_i, \phi) = \int_{n_i}^{\mu_i} q(s, n_i, \phi) ds = \int_{n_i}^{\mu_i} \frac{n_i - s}{\phi s} ds, \quad (6)$$

and the quasi-likelihood function for all sample is the summation of quasi-likelihood function for each observation as follows:

$$Q(\boldsymbol{\mu}, \mathbf{n}, \phi) = \sum_{i=1}^m Q(\mu_i, n_i, \phi) = \sum_{i=1}^m \int_{n_i}^{\mu_i} \frac{n_i - s}{\phi s} ds. \quad (7)$$

The estimated parameters will try to maximize the value of  $Q(\boldsymbol{\mu}, \mathbf{n}, \phi)$  and the estimation equation is in the following system of equations:

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = \sum_{i=1}^m q(\mu_i, n_i, \phi) \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^m \left( \frac{n_i - \mu_i}{\phi \mu_i} \right) \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = 0, \quad (8)$$

which is equivalent to

$$\sum_{i=1}^m (n_i - \mu_i) \mathbf{x}_i = 0. \quad (9)$$

And, in terms of regression parameters  $\boldsymbol{\beta}$  and exposure variable  $t$ , the system of equations is also as follows:

$$\sum_{i=1}^m [n_i - t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})] \mathbf{x}_i = 0. \quad (10)$$

In addition, the dispersion parameter  $\phi$  can be estimated by the Pearson estimator in the following equation:

$$\hat{\phi} = \frac{1}{m - p} \sum_{i=1}^m \frac{(n_i - \hat{\mu}_i)^2}{\hat{\mu}_i}. \quad (11)$$

Therefore, the Accident Hazard Index for road segment  $i$  can be formulated in the following form in this study:

$$\begin{aligned} \Lambda_i &= E \left[ \frac{w N_i^I + (1 - w) N_i^{II}}{t_i} \right] \\ &= w E \left( \frac{N_i^I}{t} \right) + (1 - w) E \left( \frac{N_i^{II}}{t} \right) \\ &= w \exp(\mathbf{x}_i^T \boldsymbol{\beta}) + (1 - w) \exp(\mathbf{x}_i^T \boldsymbol{\delta}), \end{aligned} \quad (12)$$

where  $\boldsymbol{\beta}$  is the vector of regression parameters for fatal-injury accidents and  $\boldsymbol{\delta}$  is the corresponding parameters for property-damage-only accidents.

### 3. Results and Discussion

The model is constructed by the function of generalized linear models in  $R$ . Based on the above model specification, parameters of  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$  are estimated by the quasi-Poisson model and the results are presented in Table 2, which includes only statistically significant effects (variable selection is based on backward elimination with a 0.05 significance level to stay). Interestingly, the individual impact of each covariate is largely consistent between fatal-injury accidents and property-damage-only accidents with a slight exception that several variables do not show evident impact on  $\boldsymbol{\beta}$  but on  $\boldsymbol{\delta}$ .

Road segments show different risk levels between rural and urban locations in terms of the crash rate of fatal-injury accidents. The model indicates that rural road segments are more likely to be involved in fatal-injury accidents. Even the effect of rural location is controversial [17], it is also worth to mention that one possible reason is that the driving speed is usually high and hence more likely to involve more serious

TABLE 2: Estimation results for the quasi-Poisson models.

Coefficients for parameters	Fatal-injury ( $\beta$ )		Property-damage-only ( $\delta$ )	
	Coefficient	<i>t</i> stat.	Coefficient	<i>t</i> stat.
Explanatory variables				
Intercept	-2.533	-17.383	-0.713	-3.839
Road segment in rural areas*				
Road segment in urban areas	-0.393	-2.844	#	#
Intersection density	0.100	5.922	0.130	11.310
Ownership at state or federal levels*				
Ownership at other levels	0.275	1.974	0.441	3.991
Pavement condition rating	#	#	-0.100	-2.141
Percent average daily trucks	-0.083	-4.305	-0.085	-5.071
Dispersion parameter				
$\phi$	3.096		16.415	

# Indicates that the coefficient is statistically insignificant.

\*Reference categories.

accidents like fatal or injury ones. Several previous studies also indicate that in rural areas there are higher death rates compared to urban areas due to excessive speeding [18] and special rural driving cultures [19].

Intersection density is defined by the number of intersections along the road segments per mile. The positive coefficient shows that intersection density is an unsafe factor that may introduce more chances for both fatal-injury and property-damage-only accidents to occur. Even though the samples of accidents occurred at intersections have been excluded in this study, the remaining accidents that occurred at the road segments are also affected by these intersections and are possibly due to the complicated upstream or downstream traffic flows near intersections. Specifically, the increased demand of waving and lane change actions when the vehicle approaches or leaves intersections will lead to complicated traffic flow situations as well as confictions between vehicles and hence may contribute to more risk of accidents.

As mentioned, positive coefficient is an indication of high risk for accidents and another unsafe factor is the ownership of road segments. This study will distinguish the ownership of road segments in terms of state/federal roads or roads owned by other levels of governments. Therefore, the coefficient indicates that road segments under the ownership of state or federal level will produce small AHI for accidents with all other variables being fixed. It is plausible that state or federal roads may receive better considerations from the road design to traffic operation and gain safer conditions than roads owned by town or municipal governments.

Presence of trucks is not an unsafe factor for road safety evaluation as the truck drivers are well trained [20] and hence more professional and cautious during driving than the drivers of passenger vehicles. The coefficient on the percent average daily truck is negative which means the appearance of truck is a safe factor.

Annual average daily traffic consistently associates with the occurrence of all types of accidents and the coefficients indicate that road segments with larger AADT may lead to

fewer risks. One of the possible reasons is that the operation speed is usually low for large AADT and may provide safer environment for driving. In addition, better pavement condition will also lead to fewer risks by negatively affecting the occurrence of property-damage-only accidents.

The dispersion parameters  $\phi$  are estimated by the Pearson estimator and a value greater than one indicates that overdispersion exists for the count data. For fatal-injury accident and property-damage-only accidents, the dispersion parameters are estimated and 3.096 and 16.415 are the results, respectively. Therefore, the count data is overdispersed in this analysis and the quasi-Poisson is appropriate under such consideration.

With the estimated coefficients of  $\beta$  and  $\delta$ , the risk can be estimated by its mean expression. The harmfulness weight  $w$  is used to reflect magnitude of losses that resulted from fatal-injury accidents relative to the property-damage-only accidents. So naturally,  $w$  should be greater than 0.5 and there could be many approaches on choosing a particular  $w$ . In the simplest way, it can be determined by past experiences [6] or the subjective impression on the harmful levels of fatal-injury accidents over property-damage-only accidents. Another criterion for determining  $w$  could use the relative ratio between averaged insurance claimed value of fatal-injury accidents and property-damage-only accidents. For the safety evaluation purposes, the ranks of risks of all road segments are sufficient instead of the absolute numerical values of risks. Figure 3 therefore plots the risk ranks between two different choices of  $w$  to indicate the variation on the evaluated risk levels with respect to the choice of  $w$ . Specifically,  $w = 0.5$  is used as the base case, which represents the equal importance of the two types of accidents in evaluation, to compare with several other choices of  $w$ .

The correlative relationships between two resulted ranks under difference choice of  $w$  can reflect the influence of the value of  $w$  on risk ranks of all road segments. In Figure 3, an off-diagonal line indicates that the risk ranks are consistent between two choices of  $w$ , whereas the fact that more points away from this line means more variations of risk ranks under different  $w$ . Therefore, it is clear that the risk ranks of

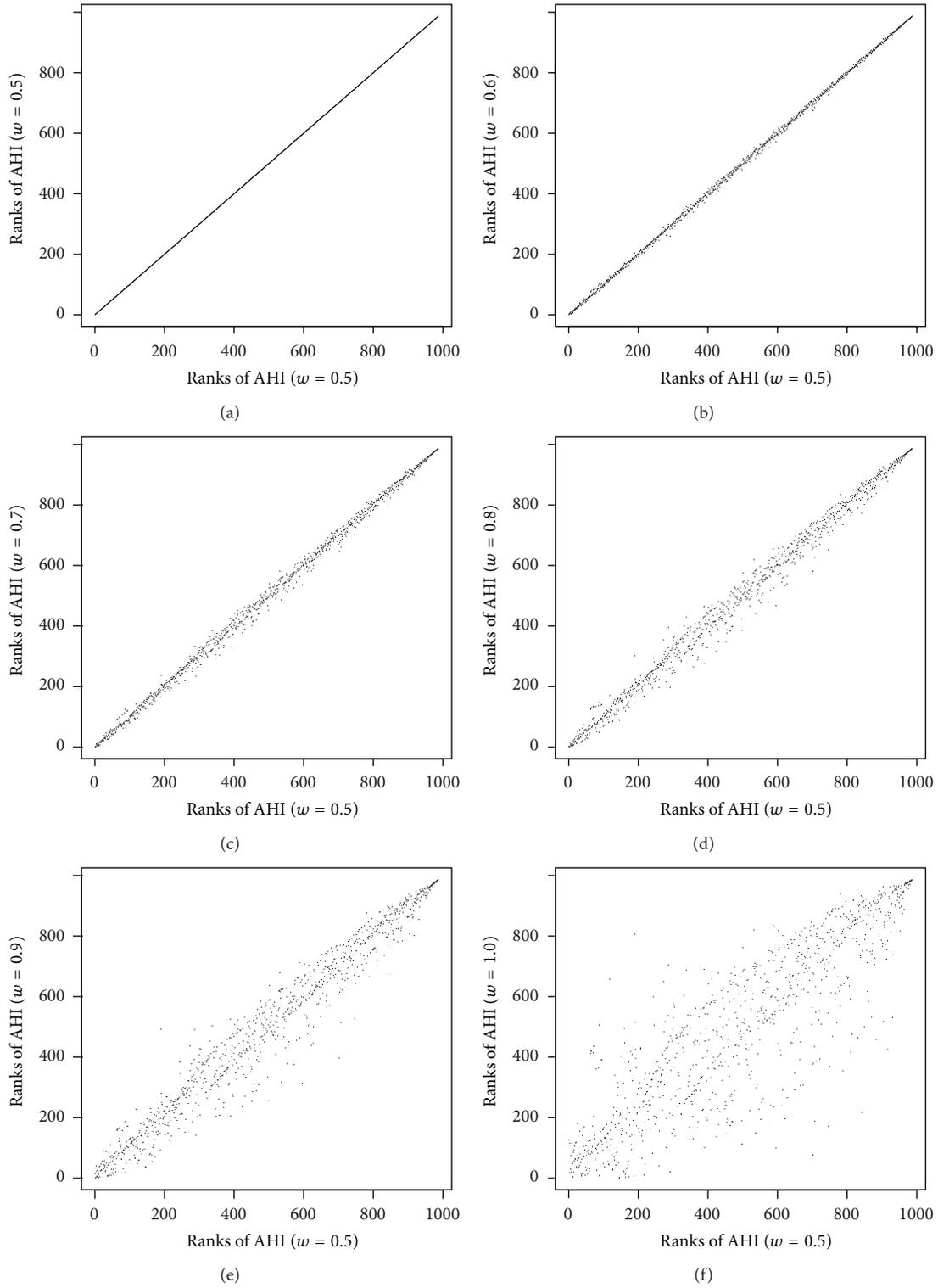


FIGURE 3: Scatter plots of AHI ranks between  $w = 0.5$  and other values of  $w$ .

AHI exhibit evident discrepancy if  $w$  is close to one and the dramatic discrepancy occurred when  $w$  is greater than 0.8. Thus, the harmfulness of accidents actually could be better considered for the selection of  $w > 0.8$  as illustrated in the plot.

#### 4. Summary and Conclusions

At the planning level, all transportation system characteristics and road traffic characteristics are important indicators which may in turn influence the roadway accident risks. The whole framework of the safety analysis at the planning level requires connections between transportation planning outputs and the accident risk evaluation criteria. Correspondingly, some of the explanatory variables in the regression models of crash counts will be used as the bridge between transportation planning and safety evaluations of planned roadway. The contribution of this paper on transportation safety planning is the developed risk evaluation models which are important preliminary works for the safety analysis at the planning level even though it is not a direct study towards safety planning.

Towards this end, this study provides a statistical analysis on a comprehensive measure which is also called the Accident Hazard Index (AHI) on accident risks. AHI is suggested as a compound value of accident frequency and corresponding harmfulness of each accident. In order to consider the overdispersed nature of the accident data, a quasi-Poisson model is proposed to connect the accident rate to several key explanatory variables. The data is integrated from several sources through the GIS platform and an clear procedure for data processing is also presented as an example for similar studies. With the aggregated accident counts on road segments, the regression model is estimated and several variables are found to have significant impact on the estimation of accident risks. For example, the intersection density has negative contributions for reducing risks, whereas AADT affects it in the opposite direction. Besides, the weight on fatal-injury accident also affects the estimated AHI and the influence can be illustrated by the changes of ranks of AHI in terms of changes of weight and the plot indicates that a value of  $w > 0.8$  is suggested for the consideration of harmfulness of fatal or injury accidents.

This work is believed to be an important first step toward a comprehensive risk analysis of traffic accidents. In addition, there are several important avenues for further research. First, it is necessary to find regression models for the accident risks as a whole such that the relationship between different types of accidents can be considered. Second, the nature of excessive zero count of accidents would be taken into account in the model as an important supplement for the traditional quasi-likelihood models.

#### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

#### Acknowledgments

This research was supported by the National Natural Science Foundation of China (no. 51208032 and no. 71210001), Fundamental Research Funds for the Central Universities (nos. 2013JBM041 and 2013JBM008), and “863” Research Project (2011AA110303).

#### References

- [1] P. C. Anastasopoulos, A. P. Tarko, and F. L. Mannering, “Tobit analysis of vehicle accident rates on interstate highways,” *Accident Analysis and Prevention*, vol. 40, no. 2, pp. 768–775, 2008.
- [2] D. Lord and F. L. Mannering, “The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives,” *Transportation Research A*, vol. 44, no. 5, pp. 291–305, 2010.
- [3] A. Tarko and M. Tracz, “Accident prediction models for signalized crosswalks,” *Safety Science*, vol. 19, no. 2-3, pp. 109–118, 1995.
- [4] S. P. Miaou, “The relationship between truck accidents and geometric design of road sections: poisson versus negative binomial regressions,” *Accident Analysis and Prevention*, vol. 26, no. 4, pp. 471–482, 1994.
- [5] D. Lord, S. P. Washington, and J. N. Ivan, “Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory,” *Accident Analysis and Prevention*, vol. 37, no. 1, pp. 35–46, 2005.
- [6] F. Holt and Ullevig, “2035 Regional transportation plan update, safety analysis-procedural overview,” Prepared for Denver Regional Council of Governments, 2008.
- [7] J. Ma and K. M. Kockelman, “Bayesian multivariate poisson regression for models of injury count, by severity,” *Transportation Research Record*, vol. 1950, pp. 24–34, 2006.
- [8] P. C. Anastasopoulos, V. N. Shankar, J. E. Haddock, and F. L. Mannering, “A multivariate tobit analysis of highway accident-injury-severity rates,” *Accident Analysis and Prevention*, vol. 45, pp. 110–119, 2012.
- [9] Y.-C. Chiou and C. Fu, “Modeling crash frequency and severity using multinomial-generalized poisson model with error components,” *Accident Analysis and Prevention*, vol. 50, pp. 73–82, 2013.
- [10] R. W. M. Wedderburn, “Quasi likelihood functions, generalized linear models, and the Gauss Newton method,” *Biometrika*, vol. 61, no. 3, pp. 439–447, 1974.
- [11] P. McCullagh, “Quasi-likelihood functions,” *Annals of Statistics*, vol. 11, no. 1, pp. 59–67, 1983.
- [12] T. A. Severini and J. G. Staniswalis, “Quasi-likelihood estimation in semiparametric models,” *Journal of the American Statistical Association*, vol. 89, no. 426, pp. 501–511, 1994.
- [13] J. M. ver Hoef and P. L. Boveng, “Quasi-poisson versus negative binomial regression: how should we model overdispersed count data?” *Ecology*, vol. 88, no. 11, pp. 2766–2772, 2007.
- [14] B. P. Y. Loo, “Validating crash locations for quantitative spatial analysis: a GIS-based approach,” *Accident Analysis and Prevention*, vol. 38, no. 5, pp. 879–886, 2006.
- [15] T. Steenberghen, T. Dufays, I. Thomas, and B. Flahaut, “Intra-urban location and clustering of road accidents using GIS: a Belgian example,” *International Journal of Geographical Information Science*, vol. 18, no. 2, pp. 169–181, 2004.

- [16] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman and Hall, New York, NY, USA, 2nd edition, 1989.
- [17] X. Yan, B. Wang, M. An, and C. Zhang, "Distinguishing between rural and urban road segment traffic safety based on zero-inflated negative binomial regression models," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 789140, 11 pages, 2012.
- [18] R. P. Gonzalez, G. R. Cummings, H. A. Phelan, S. Harlin, M. Mulekar, and C. B. Rodning, "Increased rural vehicular mortality rates: roadways with higher speed limits or excessive vehicular speed?" *The Journal of Trauma*, vol. 63, no. 6, pp. 1360–1363, 2007.
- [19] M. E. Rakauskas, N. J. Ward, and S. G. Gerberich, "Identification of differences between rural and urban safety cultures," *Accident Analysis and Prevention*, vol. 41, no. 5, pp. 931–937, 2009.
- [20] X. Zhu and S. Srinivasan, "A comprehensive analysis of factors influencing the injury severity of large-truck crashes," *Accident Analysis and Prevention*, vol. 43, no. 1, pp. 49–57, 2011.

## Research Article

# Pedestrian Detection and Tracking for Counting Applications in Metro Station

**Chen Yan-yan, Chen Ning, Zhou Yu-yang, Wu Ke-han, and Zhang Wei-wei**

*Beijing University of Technology, Chaoyang District Beijing, China*

Correspondence should be addressed to Chen Ning; chenningscut@126.com

Received 9 November 2013; Revised 12 January 2014; Accepted 19 January 2014; Published 27 February 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Chen Yan-yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A pedestrian counting method based on Haar-like detection and template-matching algorithm is presented. The aim of the method is to count pedestrians that are in a metro station automatically using video surveillance camera. The most challenging problem is to count pedestrians accurately in the case of not changing the position of the surveillance camera, because the view that surveillance camera uses in a metro station is always short-shot and nondirect downward view. In this view, traditional methods find it difficult to count pedestrians accurately. Hence, we propose this novel method. In addition, in order to improve counting accuracy more, we present a method to set the parameter value with a threshold-curve instead of a fixed threshold. The results of experiments show the high accuracy of our method.

## 1. Introduction

Metro as a mode of passenger transport with the characters of speed and comfort has been favored by the public in many cities. More and more urban residents like to choose metro as their transportation tool. Because of the large number of passengers that take metro, many cities' metro stations have often crowded especially in some rush hours. If the crowded situation cannot be evacuated promptly, it will impact the metro station's operation and management. What is more, a stampede will very likely occur if the crowded situation lasts for long or becomes severe. Therefore, to avoid the congestion and prevent accidents, metro station staff needs to get passenger flow information in real time so as to evacuate passengers promptly. Pedestrian counting information is the most important among these passenger flow information.

However, at present, the main way to get the pedestrian counting information in a metro station is to estimate the passenger flow volume from surveillance video by monitoring personnel. This way is simple and easy to implement, but it has some disadvantages that are difficult to overcome. Firstly, because of getting the information in manual collection, the pedestrian counting information that metro station staff can get is not quantitative data but qualitative data. So the information is inaccurate and easily affected by subjective factors

of monitoring personnel. Secondly, monitoring personnel is difficult to provide the metro station staff with full and correct information when monitoring points are added.

Getting pedestrian counting information in automatic collection has unparalleled advantages compared with getting the information in a manual way. These advantages are as follows. First, the automatic way can improve the precision of pedestrian counting data and turn the qualitative data into the quantitative data. Therefore, this way can provide accurate and real-time data that can help metro station staff to analyse the situation of passenger flow and handle anomalous event. Second, this way can decrease workload of the monitoring personnel and reduce influence of subjective factors.

The way to count pedestrians in a metro station automatically based on computer vision has a huge cost advantage compared to the other ways, because it can use the existing surveillance video equipment which is common in a metro station. However, the computer vision way is not mature due to the factors of complicated occlusions and cluttered scenes. For that reason, this paper studies how to count pedestrians that are in a metro station accurately using related algorithm of image processing and pattern recognition. This paper's purpose is to present a method to count pedestrians in a metro station accurately and in real time without changing the position of the existing surveillance camera.

The method presented in this paper consists of three parts. First part is to detect pedestrian using Haar-feature and MHI algorithm and then to track pedestrian using template-matching algorithm in the second part. Finally the method of threshold-curve is proposed to improve the counting accuracy.

## 2. Related Work

Recently, many image-processing based methods of counting people were proposed. As using in different scenes, these methods are quite different.

Mittal et al. have proposed a traffic management system using both audio and video data [1]. Hou and Pang have developed an effective method for estimating the number of people in a low-resolution image with complicated scenes in real time [2]. Chan and Vasconcelos have presented an approach to the problem of estimating the size of inhomogeneous crowds which are composed of pedestrians that travel in different directions [3]. Xiong et al. have proposed a potential energy-based model to estimate the number of people in public scenes as the fundamental research to detect the abnormal crowd behavior [4]. Sacchi et al. have exploited the image-processing tools for moving-object detection and classification in the context of an actual application involving the remote monitoring of a tourist site [5]. Schofield et al. have designed a system to distinguish between parts of the background scene and nonbackground objects (people) [6]. Hashimoto et al. have developed a people-counting system with human information sensors [7]. Amin et al. have presented a system for counting people in a scene using a combination of low cost, low-resolution visual and infrared cameras [8]. Huang and Chow have described a people-counting system using hybrid RBF neural network [9]. Schofield et al. have described a method for counting the number of people in any predefined scene using RAM-based neural network classifiers [10]. Kopaczewski et al. have presented an algorithm for people counting in crowded scenes based on the idea of virtual gate which uses optical flow method [11]. Vicente et al. have shown the algorithm implementation for a field-programmable gate array- (FPGA-) based design for people counting using a low-level head-detection method [12]. Conte et al. have presented a novel method by establishing a mapping between some scene feature and the number of people to provide an estimate of people count [13].

Most of the previous works can be classified into two classes by camera view. One captures the video clips by long-shot and nondirect downward view camera [1–5] and the other by short-shot and direct downward view camera [6–13] (see Figure 1). The image-processing method of the first class is the first to detect and track the foreground group by body shape feature and then segment the group into individuals for counting. And the image-processing method in the second class is the first to detect foreground group and segment it into individuals by head-shoulder shape feature and then track the individuals for counting. The second class method always has lower computational complexity than the first class

because it has no need to process occlusions. However, the view of the second class method is smaller than the first class method. Both of the two class methods can make good performances in counting people accurately by their suitable views, but they are not suitable for counting pedestrians in a metro station as its camera is always set in short-shot and nondirect downward view (see Figure 2). In this view, it is difficult to segment group into individuals because of the severe body shape deformation. To solve this problem, this paper aims to propose a novel method using Haar-like feature to detect pedestrians' head and using template-matching approach to track them for counting. The reason why we select this method is that head of the pedestrian is seldom occluded in this view. Although this method is easy to make error detected, the approach of threshold-curve can compensate it.

## 3. Detecting and Tracking Pedestrian

*3.1. Haar-Like Based Pedestrian Detection.* In this paper, pedestrian is recognized and detected using Haar-like featured detector and AdaBoost classifier. This detection process has two steps—training and recognition. The flowchart of this process is shown in Figure 3.

The first step is to train samples. First of this step, we use the mouse to cut head samples from video clips by Pedestrian Detection Software (see Figure 4). In this paper, a total of 4324 samples are collected which are composed of 1000 pedestrian head positive samples and 3324 negative samples. Each positive sample is resized to  $24 * 24$  pixels and the negative samples are collected from the Internet including images of mountain, river, cartoon, and animal. After preparing this sample set, a classifier is trained to judge whether an object is a head or not. At first glance, samples' raw pixel value is an ideal feature to train a classifier, but the challenge of calculation complexity makes it difficult to take into use. Thus, the Haar-like feature instead of raw pixel value feature has been chosen. Each Haar-like feature is composed of two or three “black” and “white” rectangles joining together—these rectangles can be up-right or rotated by 45 degrees (see Figure 5). The Haar-like features value is calculated as a weighted sum of two components: the pixel gray level values sum over the black rectangle and the sum over the whole feature area (all black and white areas) [14]. An image of  $24 * 24$  contains millions of Haar-like features, so an effective feature selecting method is necessary. AdaBoost has been proven to be an effective method in selecting classifiers from millions of weak classifiers. So the AdaBoost algorithm has been selected as the learning algorithm to train the cascade classifier. After training, a pedestrian head classifier has been got.

The second step is to recognize pedestrian by pedestrian head classifier. In order to improve the speed of recognition, the region of interest (ROI) is set in the video clips. In ROI, the head classifier is adopted to recognize pedestrian in each frame (see Figure 6). The range between upper-redline and lower-redline is ROI. Two pedestrian heads are detected and highlighted by green rectangles in ROI.

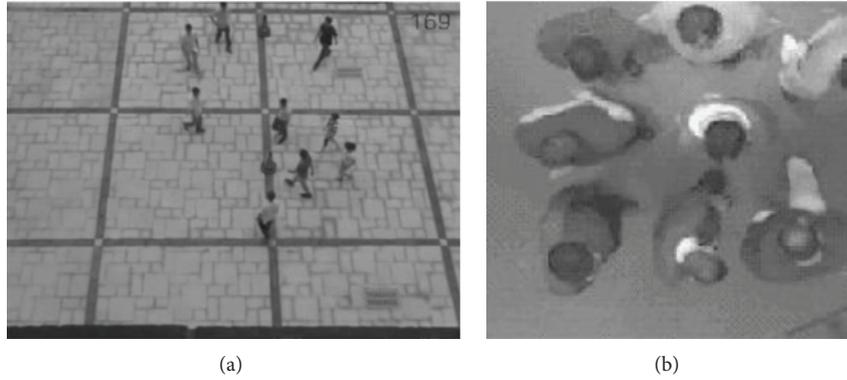


FIGURE 1: (a) Long-shot and nondirect downward view. (b) Short-shot and direct downward view.



FIGURE 2: Short-shot and nondirect downward view.



FIGURE 4: Pedestrian Detection Software.

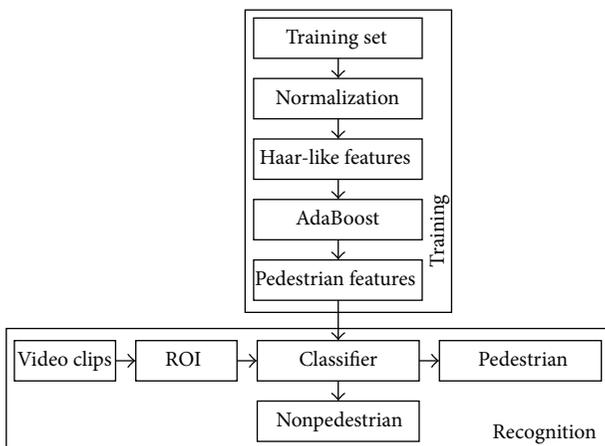


FIGURE 3: Flowchart of Haar-like based pedestrian detection.

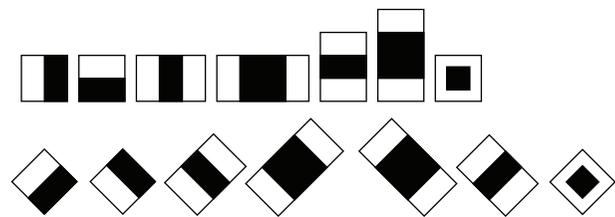


FIGURE 5: Haar-like features prototypes.



FIGURE 6: Recognizing pedestrian head in frame image.

3.2. *MHI Based Orientation Detection.* To represent how motion pedestrian works a motion history image (MHI) is formed. In an MHI,  $H_t$  pixel intensity is a function of the temporal history of motion at that point. For the results



FIGURE 7: (a) Original ROI-image. (b) MHI ROI-image.

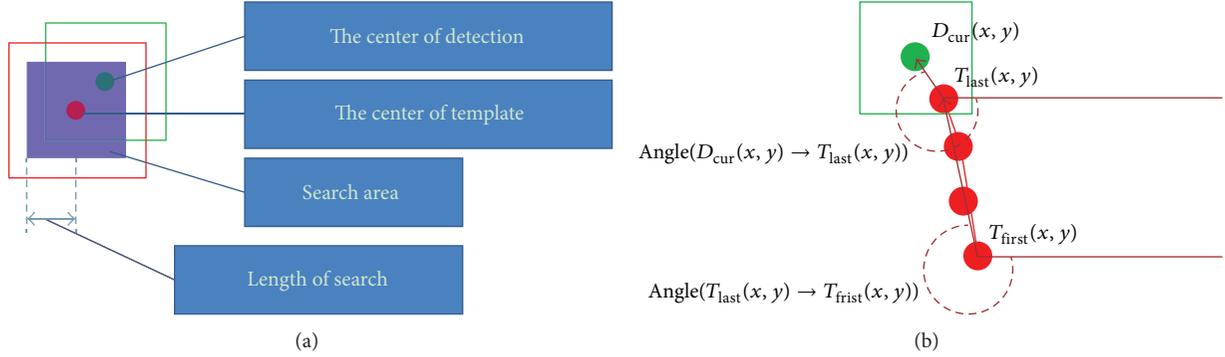


FIGURE 8: (a) Criterion 1. (b) Criterion 2.

presented here a simple replacement and decay operator is used:

$$H_r(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_r(x, y, t-1) - 1) & \text{otherwise.} \end{cases} \quad (1)$$

The result is a scalar-valued image compared with which moving pixels are brighter [15]. Examples of MHI are presented in Figure 7(b). Green circle represents the orientation of this pedestrian.

**3.3. Template-Matching Based Pedestrian Tracking.** Object association is of great importance to the application of multiple targets tracking. In order to get the associations between a single recognized rectangle and a specific template, the method of template-matching is adopted. In other words, we use template-matching algorithm to judge whether a single recognized rectangle matches to a specific template [16]. If the association between one recognized rectangle and one template satisfies the following two criteria, this recognized rectangle is considered as matching to this template.

(1) The center of recognized rectangle is located in the searching area of the template (see Figure 8(a)). The formula is listed as follows:

$$\begin{aligned} & \text{Template}(x) - \text{Length}_{\text{Search}} \\ & < \text{Detect}(x) < \text{Template}(x) + \text{Length}_{\text{Search}}, \\ & \text{Template}(y) - \text{Length}_{\text{Search}} \\ & < \text{Detect}(y) < \text{Template}(y) + \text{Length}_{\text{Search}}. \end{aligned} \quad (2)$$

In formula (2), Template  $(x)$  stands for  $x$  position of template in frame image, and Detect  $(x)$  stands for  $x$  position of detection rectangle in frame image, and so on for Template  $(y)$  and Detect  $(y)$ . Search represents the searching threshold of the template.

(2) The orientation of detection rectangle keeps the pace with the deviation angle of the template (see Figure 7(b)). The formula is presented as follows:

$$\begin{aligned} & \text{Angle}(D_{\text{cur}}(x, y) \longrightarrow T_{\text{last}}(x, y)) \\ & > \text{Angle}(T_{\text{last}}(x, y) \longrightarrow T_{\text{first}}(x, y)) - \text{AT}, \\ & \text{Angle}(D_{\text{cur}}(x, y) \longrightarrow T_{\text{last}}(x, y)) \\ & < \text{Angle}(T_{\text{last}}(x, y) \longrightarrow T_{\text{first}}(x, y)) + \text{AT}. \end{aligned} \quad (3)$$

In formula (3),  $D$  refers to the detection rectangle,  $T$  refers to the template, cur stands for the current frame, first stands for the first frame of the template, last stands for the last frame of the template, AT stands for the threshold of the angle, and  $x$  refers to  $x$  position and so does  $y$ .

If one detection rectangle matches with one template, this template is updated as in Table 1.

In order to make the template's position updated with a fast speed while the template's size with a slow speed, the specific values ( $\alpha_1 = 0.95$ ,  $\alpha_2 = 0.35$ ) are set in this paper.

If one detection rectangle cannot match each template in template list, a new template is created and added to template list. This new template is initialized as in Table 2.

While a template is unable to match with a new detection rectangle for a moment (5 frames in this paper), it is feasible to delete this template from the template list so as to judge whether or not this template is a pedestrian. In this paper,

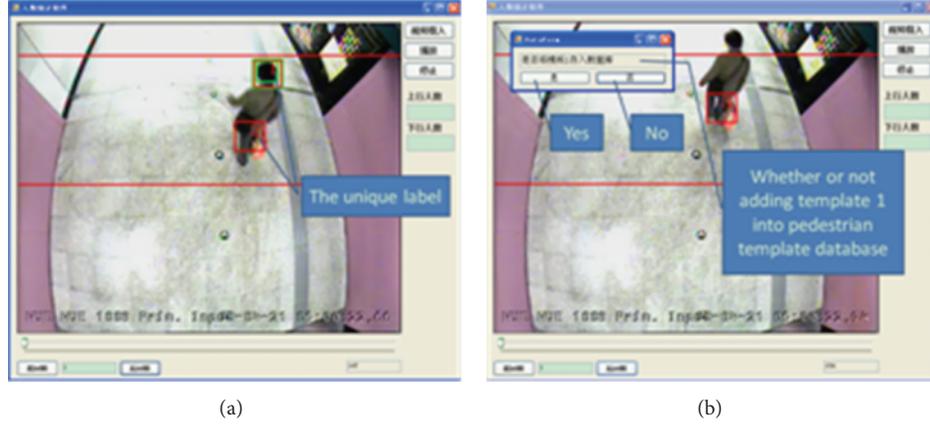


FIGURE 9: Data collection software.

TABLE 1: Template update.

Character	Explanation	Update criterion	Update method
$x$	$x$ position of center	Matching	$x = x \times (1 - \alpha_1) + \alpha_1 \times x_{\text{detection}}$
$y$	$y$ position of center	Matching	$y = y \times (1 - \alpha_1) + \alpha_1 \times y_{\text{detection}}$
$dx$	MHI accumulated value of $x$ direction	Matching	$dx+ = dx_{\text{detection}}$
$dy$	MHI accumulated value of $y$ direction	Matching	$dy+ = dy_{\text{detection}}$
Width	Template width	Matching	Width = width $\times$ (1 - $\alpha_2$ ) + $\alpha_2 \times$ width <sub>detection</sub>
Height	Template height	Matching	Height = height $\times$ (1 - $\alpha_2$ ) + $\alpha_2 \times$ height <sub>detection</sub>
Score	The number of matching with detection rectangle	Matching	Score++
Scoreperfrm	The number of matching in current frame	Matching, the end of frame	Scoreperfrm++ matching 0 the end of frame
Scoremax	The max of scoreperfrm in template	The end of each frame	Max(scoreperfrm, scoremax)
Frame	The frame of last matching	Matching	Frame = frame <sub>detection</sub>

the template is considered as a pedestrian if it meets the following conditions:

$$\begin{aligned}
 & \text{MIN\_SIZE} < \text{width} \times \text{height} < \text{MAX\_SIZE}, \\
 & \text{score} > \text{GOOD\_SCORE}, \\
 & \frac{\text{score}}{(\text{frame} - \text{startfrm} + 1)} > \text{GOOD\_SCORE\_RATIO}, \\
 & \left| \frac{dy}{(\text{frame} - \text{startfrm} + 1)} \right| > \text{GOOD\_DIR\_AVER}, \\
 & \text{score max} > \text{GOOD\_SCORE\_MAX}, \\
 & |y - \text{starty}| > \text{Y\_AXIS\_MIN}.
 \end{aligned} \tag{4}$$

In formula (4), the parameters of MIN\_SIZE and MAX\_SIZE are set for deleting some templates the size of which is too big or too small. The parameters of GOOD\_SCORE, GOOD\_SCORE\_RATIO, and GOOD\_SCORE\_MAX are set because templates created by image noise are more difficult to match than the other templates. And the parameters of GOOD\_DIR\_AVER and Y\_AXIS\_MIN are set because the

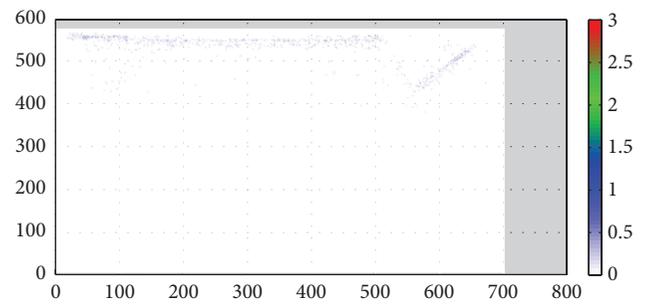


FIGURE 10: The scatter gram of the frequency of the template samples.

positions of templates produced by the noise of images change a little. The method of how to set these parameters' thresholds will be elaborated in the next section.

#### 4. Threshold-Curve Set

These parameters, which decide whether a template is regarded as a pedestrian template or not, have great influence on the accuracy of counting. Hence, in order to guarantee

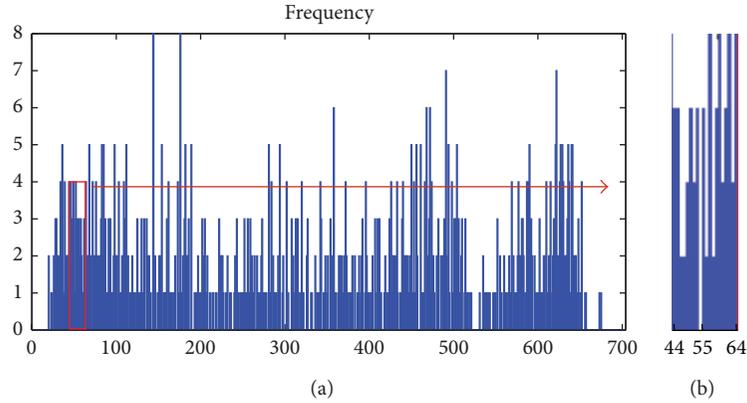


FIGURE 11: (a) The scatter gram of the frequency of the template samples in  $x$ -axis. (b) Detail with enlarged scale of (a).

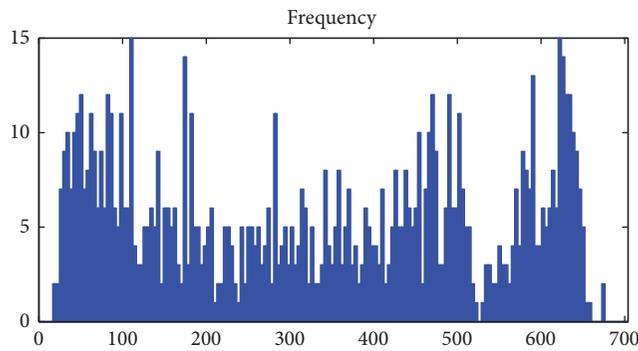


FIGURE 12: A new scatter gram of the frequency of the template samples in  $x$ -axis.

TABLE 2: Template initialization.

Character	Initialization
$x$	$x_{\text{detection}}x_{\text{detection}}$
$y$	$y_{\text{detection}}$
$dx$	$dx_{\text{detection}}$
$dy$	$dy_{\text{detection}}$
Width	Width <sub>detection</sub>
Height	Height <sub>detection</sub>
Score	1
Scoreperfrm	1
Scoremax	1
Frame	Frame <sub>detection</sub>
Startfrm	Frame <sub>detection</sub>
startx	$x_{\text{detection}}$
starty	$y_{\text{detection}}$
Label	A unique number

counting accuracy, a novel method to set these parameters' thresholds is presented. The novelty of this method is that these parameters' thresholds are set not by fixed values but by curve values. In other words, whether the template is regarded as a pedestrian or not is decided by different thresholds' values when it is in different positions of the

TABLE 3: All fields in a record.

Character	Explanation
Width, height, $dx$ , $dy$ , score, $x$ , $y$ , startx, starty, label, scoremax	These are shown in Section 2
Size	Width $\times$ height
Frames	Frame - startfrm + 1
Score ratio	Score/frames
$dx$ scorediraver	$ dx /\text{frames}$
$dy$ scorediraver	$ dy /\text{frames}$
$x$ axis	$ x - \text{startx} $
$y$ axis	$ y - \text{starty} $
$x$ axisaver	$x\text{-axis}/\text{frames}$
$y$ axisaver	$y\text{-axis}/\text{frames}$

frame image. This novel method's implementation includes two steps. The first step is data collection of the samples, and next step is to make an analysis of the data and calculate the threshold-curve.

**4.1. Data Collection.** In order to find out the different characteristics between the pedestrian template and other noise templates, a piece of software to collect pedestrian templates is designed. The piece of software is shown in Figure 9.

In this software, the display of the rectangular template and identification code are shown in the video clip in real time (see Figure 9). This software will pop up a new dialogue box named "dataForm" automatically while a template cannot match with a new detection rectangle for 5 frames. There are two buttons in this dialogue box—"Yes" and "No." The relevant characteristics of this template will be added to the pedestrian template database if the "Yes" button is pressed. In this database, each field stands for a specific feature of the template and each record represents a pedestrian template. All fields are listed in Table 3.

In this paper, 908 pedestrian templates to make a more comprehensive analysis have been collected.

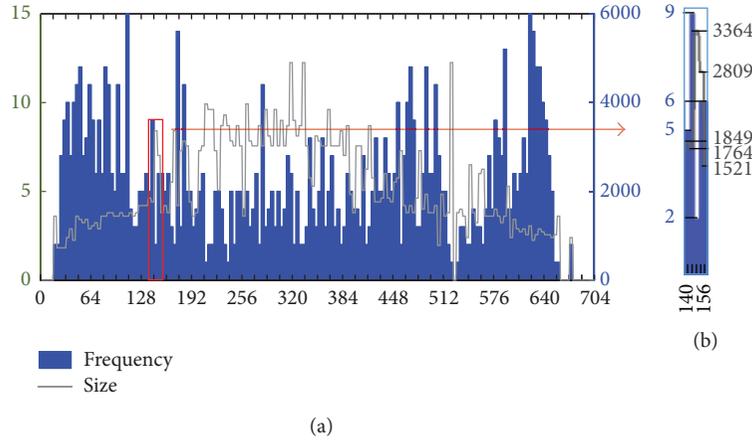


FIGURE 13: (a) MIN\_SIZE threshold figure. (b) Detail with enlarged scale of (a).

4.2. Data Analysis. From what is presented above, we judge a template by whether a pedestrian uses the parameters of MIN\_SIZE, MAX\_SIZE, GOOD\_SCORE, GOOD\_SCORE\_RATIO, GOOD\_DIR\_AVER, GOOD\_SCORE\_MAX, and Y\_AXIS\_MIN or not.

We find out that the counting error is bigger than expected if these parameters are set by fixed value. That is because template's detecting and tracking are susceptible to the surroundings. Therefore, a novel method which is to set the parameter with different values is proposed. The value is decided by the last position of the template. In other words, we want to get not a fixed value but a curve value to be the parameter's threshold, so we call it threshold-curve.

Figure 10 is the scatter gram of the frequency of the template samples. In this figure, colorful dots stand for the last position (the position where the template is deleted) of pedestrian templates which have been collected in Section 4.1, and the different colors represent the different frequency of their occurrences (red for 3 times, green for 2 times, and so on). It is observed that these dots' y-value is quite close and x-value is disperse, so we use these dots' x-value as the independent variable of the threshold-curve. To help readers better understand this new method, we will take the example of setting the threshold of MIN\_SIZE to further explain how to set these thresholds' values.

4.2.1. MIN\_SIZE Threshold-Curve. As is mentioned above, MIN\_SIZE threshold-curve has been gotten according to x-values of the template samples. The x-value is divided into 704 groups as this video resolution is 704 \* 576 pixels. Each group includes 576 pixels from (1, y<sub>n</sub>) to (576, y<sub>n</sub>). Figure 11 is the scatter gram of the frequency of the template samples in x-axis. The figures show that it is easy to make mistakes to adopt this method of classification, which are shown in Figure 11(b). In this figure, none of the x-values of the pedestrian templates comes to 55; however, it does not mean that a pedestrian will never appear at this position (55, y<sub>n</sub>). And events with low frequency may account for this error.

It is the most direct way to expand the number of the samples so as to avoid this mistake. However this solution

TABLE 4: Pixels in a big group.

$(1, y_n)$	$(1, y_{n+1})$	$(1, y_{n+2})$	$(1, y_{n+3})$
$(2, y_n)$	$(2, y_{n+1})$	$(2, y_{n+2})$	$(2, y_{n+3})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(576, y_n)$	$(576, y_{n+1})$	$(576, y_{n+2})$	$(576, y_{n+3})$

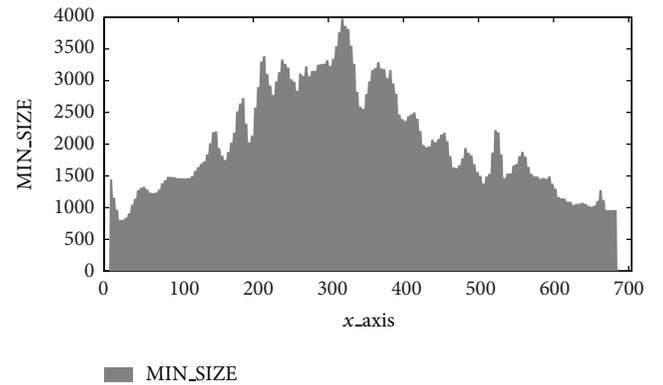


FIGURE 14: A new MIN\_SIZE threshold figure.

is not feasible because more samples would increase the intensity of manual work. The solution we adopt is to integrate four small groups into a big group, each of which includes 4 \* 576 pixels. The detail is presented in Table 4.

A new scatter gram of the frequency of the template samples in x-axis is shown after using this method (see Figure 12).

The templates in a big group form a template set, each of which has the property of size. We will get a set of the templates by this way. In order to get the threshold of the MIN\_SIZE parameter, we sort this template's size set from small to large and collect the 10~(th) percentile value of this set (see Figure 13(a)). The reason why 10~(th) percentile value is chosen instead of the smallest value lies in the principle of reducing the impact of events with low frequency.

Both the size value and the frequency distribution of the pedestrian templates have impacts on the threshold of

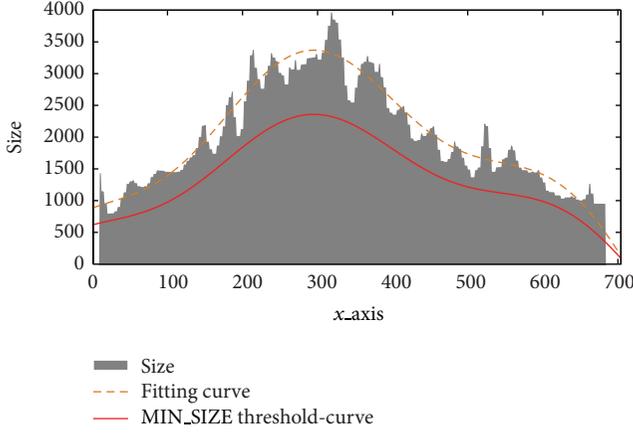


FIGURE 15: MIN\_SIZE threshold-curve figure.

TABLE 5: Set MIN\_SIZE threshold.

Group number	X-coordinate	Frequency	10~(th) percentile size value
35	137–140	5	1849
36	141–144	9	1764
<b>37</b>	<b>145–148</b>	<b>2</b>	<b>3364</b>
38	149–152	6	2809
39	153–156	6	1521

the MIN\_SIZE. For example, in Table 5, when  $x$ -coordinate value equals 145–148, we consider that the value of the MIN\_SIZE is a little larger than its actual value because its frequency is only 2 and the sizes nearby (1849, 1764, 2809, and 1521) are smaller. To guarantee the veracity of the threshold-curve, the following is used to calculate the MIN\_SIZE value of the templates:

$$f_{\text{sum}} = f_{n-2}\gamma_2 + f_{n-1}\gamma_1 + f_n + f_{n+1}\gamma_1 + f_{n+2}\gamma_2,$$

$$\text{Update } S_n = \frac{\gamma_2 f_{n-2}}{f_{\text{sum}}} s_{n-2} + \frac{\gamma_1 f_{n-1}}{f_{\text{sum}}} s_{n-1} + \frac{f_n}{f_{\text{sum}}} s_n + \frac{\gamma_1 f_{n+1}}{f_{\text{sum}}} s_{n+1} + \frac{\gamma_2 f_{n+2}}{f_{\text{sum}}} s_{n+2}. \quad (5)$$

$f_n$  stands for the frequency value of group  $n$ ;  $\gamma_1, \gamma_2$  stand for weight value (in this paper,  $\gamma_1 = 0.6, \gamma_2 = 0.4$ );  $s_n$  stands for the size value of group  $n$ ; Update  $S_n$  stands for the new size value of group  $n$ . This method to get a new MIN\_SIZE threshold figure is adopted (see Figure 14).

Now the discrete function of the thresholds of the MIN\_SIZE has been got, but it is far from enough. To make it convenient to design of the program and boost the calculating speed, we intend to get a continuous function (threshold-curve) of the MIN\_SIZE, which can be obtained by two steps.

Firstly, the discrete MIN\_SIZE threshold data is adapted to a sum of the sin function ( $f(x) = a_1 \sin(b_1 x + c_1) + a_2 \sin(b_2 x + c_2) + \dots + a_n \sin(b_n x + c_n)$ ) by fitting at least squares principle. As is known to all, the larger the  $n$ -value

becomes, the smaller the SSE-value (Sum of Squares for Error) becomes. However, it takes more time to calculate. Therefore, in order to work out a proper  $n$ -value that can make SSE-value and calculating speed in the ballpark, the value of  $n$  is increased one by one and the SSE-value of the fitting function is calculated. When SSE-value is less than  $(0.07 \times \text{MIN\_SIZE}_{\text{max}})^2 \times 704$  ( $\text{MIN\_SIZE}_{\text{max}}$  for the max value of MIN\_SIZE), the value of  $n$  can be worked out.

Secondly, to make the threshold-curve more fault-tolerant, the fitting function is multiplied by a coefficient. If the lower limit value is needed, this coefficient should be 0.7; if the upper limit value is wanted, this coefficient should be 1.3. We will get the threshold-curve of the MIN\_SIZE by this way.

The threshold-curve of the MIN\_SIZE is shown in Figure 15 and the following formula:

$$\text{MIN\_SIZE}(x) > 2056 \sin(0.004152x + 0.261) + 312.5 \sin(0.01739x + 2.857). \quad (6)$$

**4.2.2. The Other Parameters' Threshold-Curve.** The other parameters' (MAX\_SIZE, GOOD\_SCORE, GOOD\_SCORE\_RATIO, GOOD\_DIR\_AVER, GOOD\_SCORE\_MAX, and Y\_AXIS\_MIN) threshold-curve can be gotten like MIN\_SIZE. These parameters' threshold-curve is worked out as in Figures 16, 17, 18, 19, 20, and 21 and the following:

$$\begin{aligned} \text{MAX\_SIZE}(x) < & 7947 \sin(0.004436x - 0.01695) \\ & + 1033 \sin(0.009265x + 0.3037) \\ & + 521 \sin(0.02501x + 0.2283), \end{aligned} \quad (7)$$

$$\begin{aligned} \text{GOOD\_SCORE}(x) > & 8.687 \sin(0.004133x + 0.1825) \\ & + 1.727 \sin(0.01654x + 2.847) \\ & + 1.515 \sin(0.04421x - 1.328), \end{aligned} \quad (8)$$

$$\begin{aligned} \text{GOOD\_SCORE\_RATIO}(x) > & 0.7049 \sin(0.003123x + 0.6154) \\ & + 0.09464 \sin(0.01939x - 4.152), \end{aligned} \quad (9)$$

$$\begin{aligned} \text{GOOD\_DIR\_AVER}(x) > & 0.1152 \sin(0.006032x - 0.4753) \\ & + 0.07742 \sin(0.01011x + 1.219) \\ & + 0.02763 \sin(0.02273x - 0.04605) \\ & + 0.01126 \sin(0.05589x - 2.976) \\ & + 0.03896 \sin(0.08273x - 1.865) \\ & + 0.03893 \sin(0.08434x + 0.9109) \\ & + 0.01092 \sin(0.06136x - 1.576) \\ & + 0.02656 \sin(0.0232x + 2.896), \end{aligned} \quad (10)$$

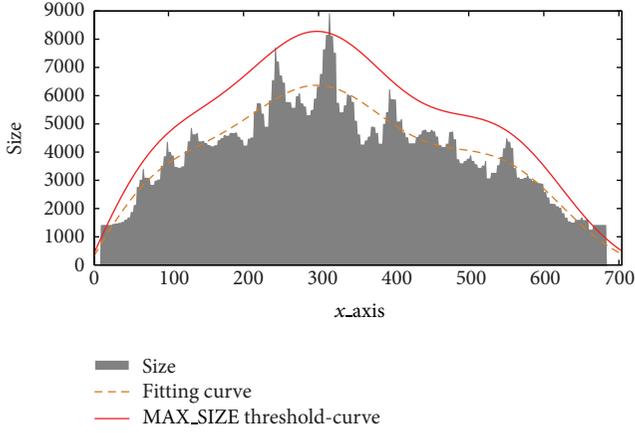


FIGURE 16: MAX\_SIZE threshold-curve figure (see (7)).

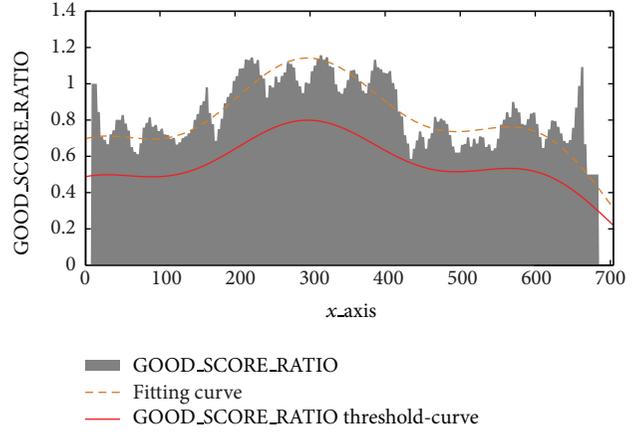


FIGURE 18: GOOD\_SCORE\_RATIO threshold-curve figure (see (9)).

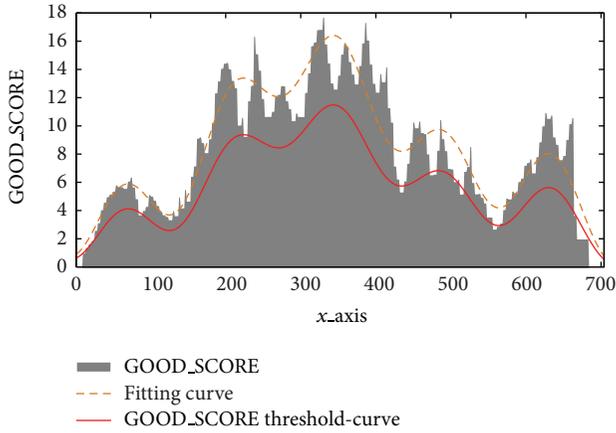


FIGURE 17: GOOD\_SCORE threshold-curve figure (see (8)).

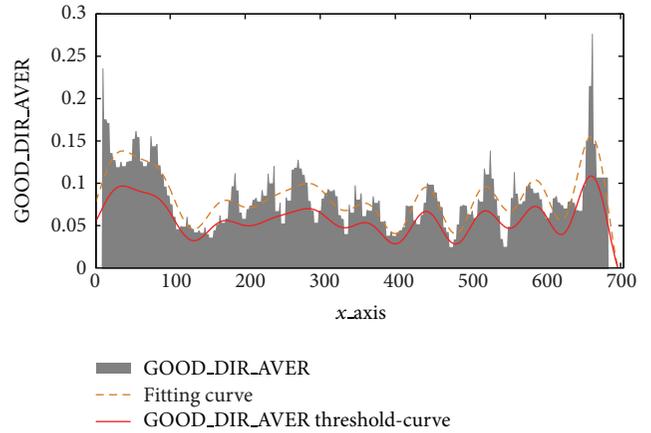


FIGURE 19: GOOD\_DIR\_AVER threshold-curve figure (see (10)).

$$\text{GOOD\_SCORE\_MAX}(x) > 1, \quad (11)$$

$$\begin{aligned} \text{Y\_AXIS\_MIN}(x) &> 56.84 \sin(0.004523x + 0.09802) \\ &+ 11.02 \sin(0.01791x + 2.443) \\ &+ 188 \sin(0.0426x - 0.538) \\ &+ 7.161 \sin(0.02761x + 2.556) \\ &+ 3.917 \sin(0.153x + 1.762) \\ &+ 3.448 \sin(0.0894x - 2.319) \\ &+ 2.787 \sin(0.1779x + 2.449) \\ &+ 181.7 \sin(0.04248x - 3.621). \end{aligned} \quad (12)$$

## 5. Results

To test the results of our method, we design a piece of software named Pedestrian Counting Software V1.0. This software has been developed by C# language and EmguCV library. EmguCV is the NET version of OpenCV which is an open source computer vision library developed by Intel that

TABLE 6: The fixed threshold we use.

Parameter	Fixed threshold
MIN_SIZE	729
MAX_SIZE	11664
GOOD_SCORE	2
GOOD_SCORE_RATIO	0.3333
GOOD_DIR_AVER	0.1945
GOOD_SCORE_MAX	1
Y_AXIS_MIN	0

includes a lot of general algorithms for image processing and computer vision.

The first experiment is to test the accuracy of counting pedestrian with fixed threshold. The value of parameters' thresholds is decided by the pedestrian template set we collect in Section 4.1. For example, the value of MIN\_SIZE's threshold is the minimum value of template set. The values we use in this paper are as in Table 6.

The accuracy results of counting pedestrian with fixed threshold are as in Table 7.

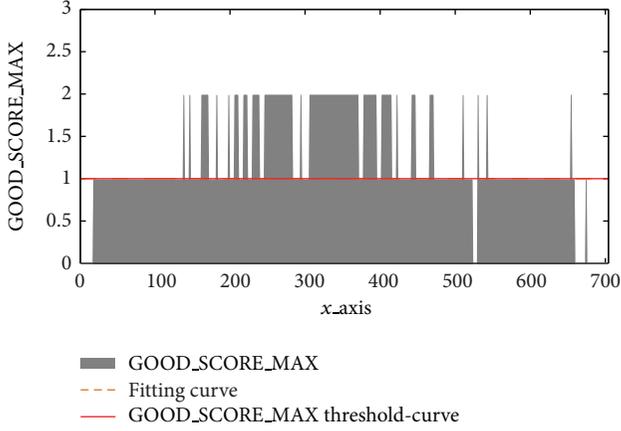


FIGURE 20: GOOD\_SCORE\_MAX threshold-curve figure (see (11)).

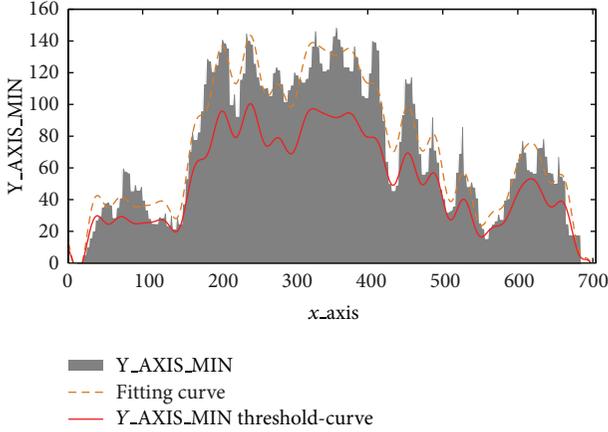


FIGURE 21: Y\_AXIS\_MIN threshold-curve figure (see (12)).

TABLE 7: Experimental results with fixed threshold.

	Up	Down
Actual data	1749	21
Missed	411	1
Mistake	91	0
Counting data	1429	20
Accuracy	81.7%	95.2%
Actual accuracy	71.3%	95.2%

And then we test the accuracy of counting pedestrian with threshold-curve. The value of parameters' thresholds we use is from Section 4.2. The accuracy results are as in Table 8.

Compared with using fixed threshold, using threshold-curve has made a performance on counting accuracy. As seen from Tables 6, 7, and 8, the method with threshold-curve improves counting accuracy by 13.9 percent and actual accuracy by 17.7 percent compared to the method with fixed threshold; hence the method using threshold-curve is much better. In addition, the accuracy is the same whether we use fixed threshold or use threshold-curve. That is because down passenger flow is very little and it is difficult to calculate

TABLE 8: Experimental results with threshold-curve.

	Up	Down
Actual data	1749	21
Missed	135	1
Mistake	58	0
Counting data	1672	20
Accuracy	95.6%	95.2%
Actual accuracy	89.0%	95.2%

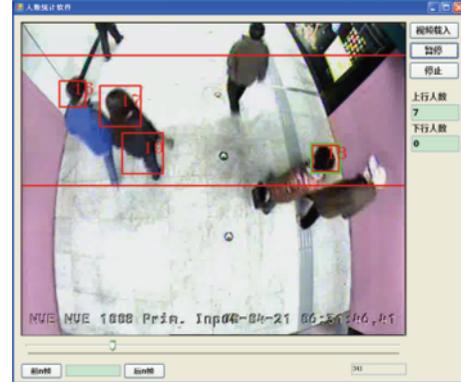


FIGURE 22: Pedestrian counting software.

threshold-curve. Therefore, we use fixed threshold instead of threshold-curve to count down passenger flow.

Although the method we present in this paper makes a good performance on counting pedestrian, counting errors like mistake and missing still exist. There are two reasons of missing counting. One is due to pedestrian wearing hats with a similar color to the surrounding. The other is pedestrian's fast speed of running. And the reason causing mistake counting is that pedestrian luggage is similar to pedestrian head.

The test video clip has been taken from Xizhimen metro station (see Figure 22). It is composed of 90000 frames, for a total amount of one hour. Its frame size is  $704 * 576$  pixels and frame rate is 25 (one frame is 40 ms). The CPU we use in experiment is Intel(R) Core(TM) 2 Duo with 3.0 GHz and the software needs about 27 ms to process one frame image. Hence, this method can meet the demands of real-time processing.

## 6. Conclusion

This paper proposes a novel method to count pedestrians in metro station based on Haar-like detection and template-matching algorithm. This new method avoids the disadvantages of other computer vision methods in short-shot and nondirect view. Thus there are the following two points between our method and the other traditional methods.

(1) To avoid counting error due to occlusions and body-shape deformation in short-shot and nondirect view, this method uses Haar-like and AdaBoost algorithms which are mature in face recognition skill to detect pedestrian.

(2) Template-matching algorithm which includes seven parameters (MIN\_SIZE, MAX\_SIZE, GOOD\_SCORE, GOOD\_SCORE\_RATIO, GOOD\_DIR\_AVER, GOOD\_SCORE\_MAX, and Y\_AXIS\_MIN) is presented in our method to track and count pedestrian. The value of these parameters is the key to judge a template whether a pedestrian or not; thus it has a great influence on the accuracy of counting pedestrian. Nevertheless, we found that the accuracy results cannot meet our demands if we set these parameters with the fixed value. Therefore, a novel method to set these parameters with the threshold-curve is proposed. In other words, a threshold function is built to every parameter so as to improve the accuracy of pedestrian counting.

The experiments prove that our method can make a good performance on counting application. And the accuracy of the method with threshold-curve is nearly ninety percent which is higher than that of the fixed threshold method's seventeen percent, so the method with threshold-curve is proved to be a better method.

The method presented in this paper is of practicable value to apply in a metro station, but there are also two disadvantages to this method. One is that a lot of workload is needed to collect template samples, so we expect to adopt some automatic collection methods to collect the samples instead of the manual collection method in the future. The other is that we must judge a template whether a pedestrian or not manually when we collect template samples; thus an automatic method to judge the template's property is expected to be proposed in the future.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This study is supported by the Science and Technology Project of Ministry of Transport: The Study on Key Technology of Transfer Information Sensing and Co-Scheduling, no. 2012-364-220-109; and Science and Technology Planning Project of Beijing Science and Technology Commission: The Study on the Mechanism of Travel Behavior and Guiding Strategy, no. Z1211000003120100.

## References

- [1] A. Mittal, A. Jain, and G. K. Agarwal, "Audio-video based people counting and security framework for traffic crossings," *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 49, no. 3, pp. 377–391, 2007.
- [2] Y. Hou and G. K. H. Pang, "People counting and human detection in a challenging situation," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 41, no. 1, pp. 24–33, 2011.
- [3] A. B. Chan and N. Vasconcelos, "Counting people with low-level features and Bayesian regression," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2160–2177, 2012.
- [4] G. Xiong, J. Cheng, X. Wu, Y. Chen, Y. Ou, and Y. Xu, "An energy model approach to people counting for abnormal crowd behavior detection," *Neurocomputing*, vol. 83, pp. 121–135, 2012.
- [5] C. Sacchi, G. Gera, L. Marcenaro, and C. S. Regazzoni, "Advanced image-processing tools for counting people in tourist site-monitoring applications," *Signal Processing*, vol. 81, no. 5, pp. 1017–1040, 2001.
- [6] A. J. Schofield, T. J. Stonham, and P. A. Mehta, "Automated people counting to aid lift control," *Automation in Construction*, vol. 6, no. 5-6, pp. 437–445, 1997.
- [7] K. Hashimoto, C. Kawaguchi, S. Matsueda, K. Morinaka, and N. Yoshiike, "People-counting system using multisensing application," *Sensors and Actuators A*, vol. 66, no. 1-3, pp. 50–55, 1998.
- [8] I. J. Amin, A. J. Taylor, F. Junejo, A. Al-Habaibeh, and R. M. Parkin, "Automated people-counting by using low-resolution infrared and visual cameras," *Measurement*, vol. 41, no. 6, pp. 589–599, 2008.
- [9] D. Huang and T. W. S. Chow, "A people-counting system using a hybrid RBF neural network," *Neural Processing Letters*, vol. 18, no. 2, pp. 97–113, 2003.
- [10] A. J. Schofield, P. A. Mehta, and T. J. Stonham, "A system for counting people in video images using neural networks to identify the background scene," *Pattern Recognition*, vol. 29, no. 8, pp. 1421–1428, 1996.
- [11] K. Kopaczewski, M. Szczodrak, A. Czyzewski, and H. Krawczyk, "A method for counting people attending large public events," *Multimedia Tools and Applications*, 2013.
- [12] A. G. Vicente, I. B. Munoz, P. J. Molina, and J. L. L. Galilea, "Embedded vision modules for tracking and counting people," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 9, pp. 3004–3011, 2009.
- [13] D. Conte, P. Foggia, G. Percannella, F. Tufano, and M. Vento, "A method for counting moving people in video surveillance videos," *Eurasip Journal on Advances in Signal Processing*, vol. 2010, Article ID 231240, 10 pages, 2010.
- [14] M. Goncalo, P. Paulo, and N. Urbano, "Vision-based pedestrian detection using Haar-like features," in *Actas do Encontro Científico Guimarães*, pp. 45–50, April 2006.
- [15] A. A. Shaikh, D. K. Kumar, and J. Gubbi, "Automatic visual speech segmentation and recognition using directional motion history images and Zernike moments," *The Visual Computer*, vol. 29, no. 10, pp. 969–982, 2013.
- [16] T. Pallejà, A. Guillamet, M. Tresanchez et al., "Implementation of a robust absolute virtual head mouse combining face detection, template matching and optical flow algorithms," *Telecommunication Systems*, vol. 52, no. 3, pp. 1479–1489, 2013.

## Research Article

# Pedestrian Guiding Signs Optimization for Airport Terminal

Jianxin Lin,<sup>1,2</sup> Rui Song,<sup>1</sup> Jifeng Dai,<sup>2</sup> and Pengpeng Jiao<sup>2</sup>

<sup>1</sup> School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

<sup>2</sup> School of Civil and Transportation Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China

Correspondence should be addressed to Jianxin Lin; linjianxin@bucea.edu.cn

Received 24 September 2013; Revised 24 December 2013; Accepted 27 December 2013; Published 25 February 2014

Academic Editor: Huimin Niu

Copyright © 2014 Jianxin Lin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The pedestrian guiding sign (PGS) is used to lead people within the transportation terminal to their directions efficiently and without boundaries. In this paper, we aim to optimize the guiding signs for people in the comprehensive transportation terminal with a mathematical model, which describes the pedestrian's reaction, judgment, and perception of the outline about the guiding signs, as well as pedestrian's moving status through self-organized characteristic behavior. Furthermore, the model also reflects the information intensity of the guiding signs within the pedestrian's visual field which is taken as the influence level score of PGS. In order to solve the model, cellular automation (CA) is employed to simulate the characteristics of the pedestrians such as crowd moving and sign selection.

## 1. Introduction

The pedestrian guiding sign (PGS) plays an important role in planning and designing of an airport terminal (AT). Generally, an AT features a complicated interior structure, and therefore the possibilities of getting lost rise substantially when pedestrians have to pick a way from multiple routes. It was found that the pedestrian had to rely on guiding signs to reach the destination or to find the right direction inside a terminal. Inside of an AT, getting the proper direction, entrance, and exit should rely on the help of PGS, which could quickly lead the pedestrian to learn about the corresponding information then gain an optimum track. Consequently, current researches heavily focus on the study of the pedestrian behaviour within AT. Capacity based models are widely used to calculate waiting time, service time, or queue length at individual facilities within the airport terminal [1–5]. On the other hand, the layout-based models are used to depict the suitability or efficiency of the planning and design of an AT [5–11].

Assessing level of service (LOS) of AT is another hot research issue in recent years. Various models, such as the fuzzy logic model [11], the binary logit model [12], and the regression analysis model [13], were applied in numerous circumstances [14]. However, limited study has been developed

to evaluate LOS within the AT according to the most critical facility [15–17], that is, the PGS.

In this research, an analysis of the characteristics of the PGS within the AT was conducted. The numerical simulation combining pedestrian features, terminal environment, and signs' location was carried out. In the description of pedestrians' herd behavior features, conditions that satisfy pedestrian self-organization behaviour are supplied. Considered for the efficiency, reliability, and the characteristics in airport terminal, the cellular automation (CA) model is used to realize the environmental condition in AT, based on which, the guiding signs location model is formulated according to the crowd conflict and congestion conditions determined via CA.

The rest of this paper is outlined as follows. Section 2 describes the issue of pedestrian guiding sign locating optimization and the application condition for the pedestrian simulation model. Section 3 builds the characteristic model of pedestrian guiding signs. Section 4 presents the self-organizing characteristic model for the pedestrian. Sections 5 and 6 discuss the influence of guiding signs on the pedestrian. Model applications are described in Section 7. Section 8 summarizes the research results and discusses future research directions.

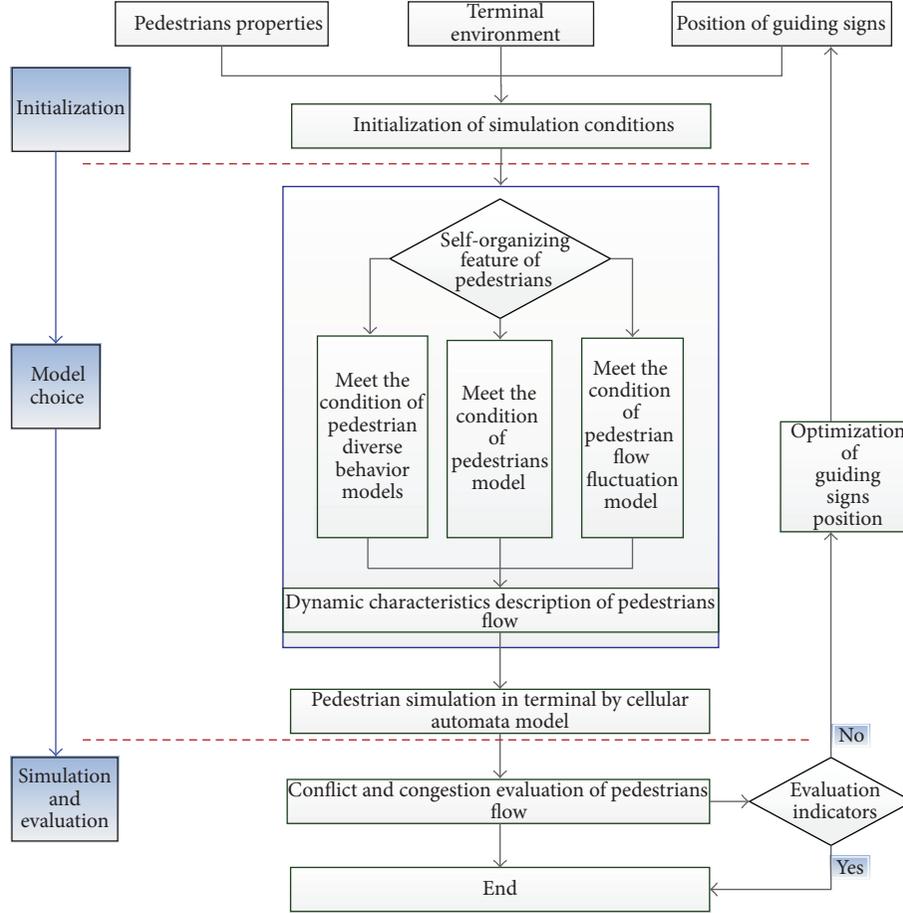


FIGURE 1: The optimization framework of pedestrian guiding sign location.

## 2. Problem Description

As showed in Figure 1, the optimization framework for setting up guiding signs within an AT is divided into 3 parts: the initiation process, the model selection process, and finally the simulation and assessment process.

There are three parts in the optimization method of PGS: initialization, model choice, and simulation and evaluation. During the initiation process, we initiate the simulation conditions, terminal environment, and guiding signs locations as numerical models. In the description of the crowd situation, the corresponding crowd simulation model will be adopted through judging the condition satisfying pedestrian self-organizing characteristic embedded in CA. Finally through the assessment of the crowd conflicts congestion resulting from simulation, we optimize the guiding signs' location.

## 3. Pedestrian Guiding Sign Characteristic Model

Generally, there are two main factors to affect the information perception of PGS: color features and contour perception. In order to describe the pedestrian cognitive behavior at PGS on color features in the simulation process, we extract

significant characteristic through PGS's color features (hue, saturation, and luminance) in terminal and get the weight of comprehensive saliency by eye movement experiment and then add the perception features above to pedestrian agent.

*3.1. Characteristic Extraction of the Outline Features of Pedestrian Guiding Sign.* Color features of PGS are used to judge the location of the signs. However, the outline features as well as the significance identification should take into account the identification of the PGS. Here we adopt the hue contrast, saturation contrast and the luminance contrast to identify the outline features of PGS.

- (1) Hue contrast:  $H_p(x, y)$  represents the hue angle of the  $P$  coordinate of PGS images; thus the hue contrast of coordinates  $(x, y)$  could be represented as

$$D_p^h(x, y) = \left| \frac{H_p(x, y) - \sum_{x \in P, y \in P} H_p(x, y) / |N_P|}{180} \right|, \quad (1)$$

where  $N_P$  denotes the pixels in the image  $P$ , that is,  $\sum_{x \in P, y \in P} H_p(x, y) / |N_P|$  is the average hue angle of pedestrian boot logo images, and  $G_p^h(x, y)$  denotes

pixel  $(x, y)$ 's hue contrast degrees, so  $G_p^h(x, y) = 1 - D_p^h(x, y)$ .

- (2) Saturation contrast:  $S_p(x, y)$  represents the saturation value of pedestrian boot logo image coordinates  $(x, y)$ , so the saturation contrast of the coordinates  $(x, y)$  is

$$D_p^s(x, y) = \left| S_p(x, y) - \sum_{x \in P, y \in P} \frac{S_p(x, y)}{|N_p|} \right|, \quad (2)$$

where  $\sum_{x \in P, y \in P} S_p(x, y)/|N_p|$  is the average saturation of the pedestrian boot logo image.  $G_p^s(x, y)$  is the saturation contrast degree of the pixel  $(x, y)$ , so that  $G_p^s(x, y) = 1 - D_p^s(x, y)$ .

- (3) Luminance contrast:  $V_p(x, y)$  represents the luminance value of the PGS image coordinates  $(x, y)$ , so that the luminance contrast of the coordinates  $(x, y)$  could be calculated as

$$D_p^v(x, y) = \left| V_p(x, y) - \sum_{x \in P, y \in P} \frac{V_p(x, y)}{|N_p|} \right|, \quad (3)$$

where  $\sum_{x \in P, y \in P} V_p(x, y)/|N_p|$  is the average saturation of the PGS images and  $G_p^v(x, y)$  is the luminance contrast degree of pixel  $(x, y)$ , and then we have  $G_p^v(x, y) = 1 - D_p^v(x, y)$ .

**3.2. PGS Contour Boundary Judgment.** When using the color contrast, saturation contrast, and luminance contrast value to carry out the extraction of the outline feature, we let  $\lambda_x$  and  $\lambda_y$  represent the search step of the flag outline feature extraction in the directions of  $x$ - and  $y$ -axes. Boundary identification conditions are set as (a)  $|G_p^h(x, y) - G_p^h(x - \lambda_x, y)| \geq G_H$  and (b)  $|G_p^s(x, y) - G_p^s(x - \lambda_x, y)| \geq G_S$ ,  $|G_p^v(x, y) - G_p^v(x - \lambda_x, y)| \geq G_V$ , where  $G_H$ ,  $G_S$  and  $G_V$  is the change threshold of the PGS's hue, saturation, and luminance boundary, respectively. The main feature extraction steps are as follows:

- (1) initialization  $x = 0$  and  $y = 0$ ;
- (2)  $x := x + \lambda_x$ ;
- (3) if  $x > X_p$ , then go to (5); otherwise continue;
- (4) after working out the hue, saturation, and luminance contrast degree value of two positions of  $(x, y)$  and  $(x - \lambda_x, y)$ , the contour boundary region could be easily determined. Meanwhile, we classify each position  $(x, y)$  into the contour boundary points set  $B$ , highlighting the significance degree in the region enclosed by four points  $(x - \lambda_x, y - \lambda_y)$ ,  $(x - \lambda_x, y)$ ,  $(x, y)$ , and  $(x - \lambda_x, y)$ , and then go to Step (2);
- (5)  $y := y + \lambda_y$ ;
- (6) if  $y > Y_p$ , end the process; otherwise  $x = 0$ , go to Step (2).

### 3.3. PGS Contour Feature Perception

- (1) We sort the outermost points of the contour boundary points set according to the order of high to low in coordinate  $Y$ . In addition, we set the  $X$  coordinates under the same  $Y$  coordinate from low to high to obtain a new boundary points sequence  $B' = (d_1(x_1, y_1), d_2(x_2, y_2), \dots, d_n(x_n, y_n))$ , where  $d_i(x_i, y_i)$   $i = 1, 2, \dots, n$  represents the points in the sequence.
- (2) Let  $\eta_b$  represent the boundary interference noise threshold of the landside guiding signs contour feature and scan the area within a radius around all points in the collection  $C$ ; all the adjacent points into region are displayed in prominent significant degree.
- (3) Use the Hough transform to identify the area after eliminating the interference above and detect the contour boundary in the image.
- (4) Different shapes of figure, such as irregular figure, equilateral triangle, and circle and rectangle, are taken as the basis of the degree of influence of the PGS on the pedestrians.

## 4. Building Pedestrian Self-Organizing Characteristic Model

**4.1. Pedestrian Distributary Behavior Model.** Crowd on the opposite direction will form an ordered stream which will not take much of the space in the opposite direction. Pedestrian behaviors in an AT mainly include decelerating and avoidance. With no distributary as well as communication with others, the pedestrian will automatically form the opposite marching stream. The number of the conflicting streams depends on the width and length of walking space, input, and output of the crowd and the waving and interference among the streams, which is the most typical example of self-organization. This kind of distributary reduces the resistance and conflicts from the contradictory stream, while the average speed and marching efficiency can be maximized. When two opposing streams meet, narrow paths could be formed at first. As time flows, pedestrian streams will become wider in order to reduce "friction" from the opposite direction which will get people on the same side to move faster and more fluently. The pedestrian stream distributary will also be autocanalized. Furthermore, the application condition for pedestrian distributary behavior model is that there are pedestrians walking oppositely in easily recognizable streams.

Figure 2 is a chart of a simplified pedestrian diversion. Let  $l_N^R$  and  $l_L^R$  represent the distances between the front and rear pedestrians, respectively, in the forward direction,  $l_N^L$  and  $l_L^L$  represent the distances between the front and rear pedestrians in the reverse direction flow,  $S_R$  and  $S_L$  represent the distance of visual impact on route selection of the forward and reverse pedestrian flows,  $\rho_R(S_R)$  and  $\rho_L(S_R)$  represent the pedestrian flow density of the forward and reverse direction in the range within  $S_R$  and  $\rho_R(S_L)$  and  $\rho_L(S_L)$  represent the

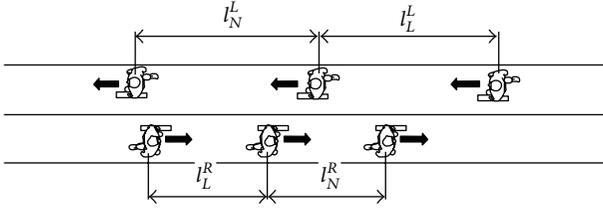


FIGURE 2: Simplified pedestrian distributary.

pedestrian flow density of the forward and reverse direction in the range within  $S_L$ . The formation and dispersion rules of the pedestrian flow are demonstrated as follows.

- (1) If  $\rho_R(S_R) \leq P_R^L$  and  $\rho_L(S_L) \leq P_L^L$ , pedestrian flows cannot be formed both in the forward and reverse direction. Pedestrians will walk freely or in a turbulent state.  $P_R^L$  denotes the lower threshold of the forward pedestrian flow.  $P_L^L$  denotes the lower threshold of the reverse pedestrian flow.
- (2) If  $\rho_R(S_R) > P_R^H$  and  $\rho_L(S_L) > P_L^H$ , orderly pedestrian flows will be formed in both forward and reverse direction, in which the mutual interference between the pedestrian is few.  $P_R^H$  denotes the upper threshold of the forward pedestrian flow.  $P_L^H$  denotes the upper threshold of the reverse pedestrian flow.
- (3) If  $\rho_R(S_R) > P_R^H$  and  $\rho_L(S_L) \leq P_L^L$ , the orderly flow of the forward pedestrian flow expands with step  $\lambda_R$ . The orderly flow of the reverse pedestrian flow decreases in step  $\lambda_L$ .
- (4) If  $\rho_R(S_R) \leq P_R^H$  and  $\rho_L(S_L) > P_L^H$ , the orderly flow of the reverse pedestrian flow expands with step  $\lambda_L$ . On the other hand, the orderly flow of the forward pedestrian flow decreases with step  $\lambda_R$ .
- (5) If  $\rho_R(S_R) \leq P_R^L$  and  $P_L^L < \rho_L(S_L) \leq P_L^H$ , the orderly flow of the reverse pedestrian flow tends to expand. The orderly flow of the forward pedestrian flow tends to decrease. Moreover, if  $\rho_L(S_L) \leq P_L^L$  and  $P_R^L < \rho_R(S_R) \leq P_R^H$ , the orderly flow of the forward pedestrian flow tends to expand. The orderly flow of the reverse pedestrian flow tends to reduce. Whether the pedestrian flow expands or decreases depends on the parameters  $l_N^R, l_L^R, l_N^L, l_L^L$ .

**4.2. Model of the Pedestrian Flow Fluctuation.** Let  $a(k, t)$  represent the accelerated velocity of the  $k$ -type pedestrian at the time  $t$ . The general acceleration formula of the  $k$ th type pedestrian at time  $t + 1$  can be expressed as  $a(k, t + 1) = \omega_1(k, p)v(k, t)^{\omega_2(k)}(\Delta v(t)^{\omega_3(k)}/\Delta x(t)^{\omega_4(k)})$ , where  $v(k, t)$  denotes the velocity of  $k$ th type pedestrian at time  $t$ .  $\Delta v(t)$  denotes the current velocity difference of the pedestrian and the front pedestrian.  $\Delta x(t)$  denotes the distance between the pedestrian and the front pedestrian.  $\omega_1(k, p)$  denotes the influence coefficient of the pedestrian's position  $p$  ( $= 1, 2, 3$  denote the head, the middle, and the end, resp.) among the

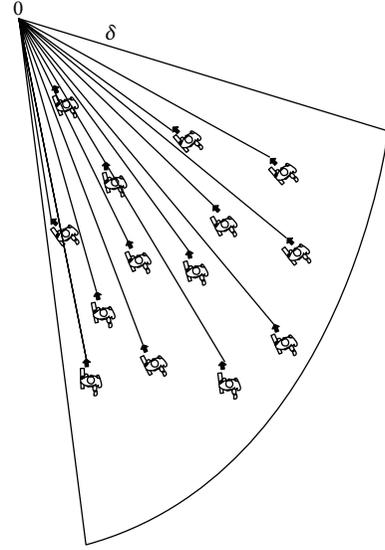


FIGURE 3: Conformity emergence chart.

group of pedestrians.  $\omega_2(k)$  represents the influence coefficient of the pedestrian's accelerated speed.  $\omega_3(k)$  represents the impact of speed difference between current pedestrian and the front pedestrian of his acceleration.  $\omega_4(k)$  represents the impact of the distance between current pedestrian and the front pedestrian of his acceleration. As a result, we can get the pedestrian's speed at the time  $t + 1$  as  $v(k, t + 1) = v(k, t) + a(k, t + 1)$ , according to which, we can easily obtain the position of the pedestrian in the next moment.

**4.3. Pedestrian Herd Behavior Model.** Conformity is the act of matching attitudes, beliefs, and behaviors to group norms when an individual acts on his own. Even if a reasonable person knows that other people are in conformity, he would participate and take similar actions. If a pedestrian is unfamiliar with the terminal environment, unclear on the routing selection, he will lose confidence in his perception and judgment, taking the majority's perception and judgment as his private grounds unconsciously, losing personal judging. In order to increase accuracy, people prefer to be with other people. Nevertheless, this kind of conformity will often cause congestion at the exit. This paper will build models on conformity in order to emulate the situation of dispersing the congestion at the exit.

**Herd Behavior Generation Condition.** To generate herd behavior, certain conditions need to be achieved in the population size and the moving direction, in a certain view field within the threshold  $\delta$ , as shown in Figure 3 the pedestrian density in the target direction of  $O$  is  $\rho$ , if  $\rho > \rho_c$ , and  $\sum_{k \in S(\delta)} v(k, t) \cdot \tan \theta_o(k, t) > V_c$  that denotes that insiders walking in this region meet the herd behavior model conditions, in which  $\rho_c$  denotes the minimum pedestrian density when herd behavior occurs,  $S(\delta)$  denotes the area of the field of view  $\delta$ , and  $v_r^o(k, t)$  denotes the speed of the pedestrian  $k$  at time  $t$ , which denotes the angle between the speed direction and the target  $O$ .

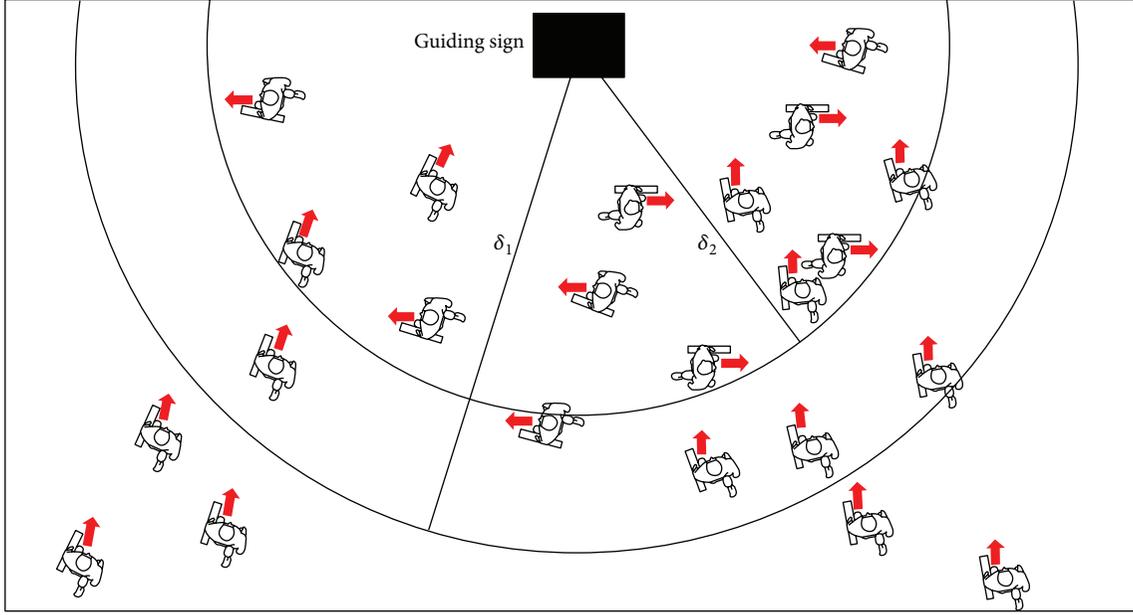


FIGURE 4: Chart of the influence that guiding signs have on pedestrians.

4.3.1. *The Speed of Herd Behavior.*  $v_e(k, t)$  represents the desired speed of the  $k$  class pedestrians at time  $t$ , then  $v_e(k, t) = [1 - p(k, t)]v_n(k, t) + p(k, t)v_{\max}(k, t)$ ,  $v_n(k, t)$  denotes the normal speed of the  $k$  class pedestrian at time  $t$ ,  $v_{\max}(k, t)$  denotes the possible maximum speed of the  $k$  class pedestrians at time  $t$ , and  $p(k, t)$  denotes the degree of urgency of the tasks of  $k$  class pedestrians at time  $t$ ; it is between 0 and 1. Under the premise of getting the desired speed, we can obtain the speed at the next time based on the current and the desired speed;  $v(k, t+1) = v(k, t) + v_e(k, t) \cdot \eta$ ,  $v(k, t)$  and  $v(k, t+1)$  denote the speed of the  $k$  class pedestrians at time  $t$  and  $t+1$ .  $\eta$  denotes the coefficient of acceleration and deceleration.

#### 4.3.2. The Direction of Herd Behavior

- (1) If  $\pi/2 < \theta_o(k, t) \leq 3\pi/2$ , the direction of pedestrian is free of the impact from the herd behavior patterns. That is,  $\theta_o(k, t+1) = \theta_o(k, t) + \omega_r$ , where  $\omega_r$  is a smaller random number.
- (2) If  $\theta_o(k, t) \leq \pi/2$  or  $\theta_o(k, t) > 3\pi/2$ , the direction of the pedestrian is in relation with both the walking speed and the density of pedestrians. The adopted rule is  $\theta_o(k, t+1) = \theta_o(k, t) + \omega(\rho(t), v(k, t))$ , where  $\rho(t)$  denotes the pedestrian density in the pedestrian's vision range at time  $t$ .  $\omega(\rho(t), v(k, t))$  denote the adjusted coefficient of the pedestrian's walking direction of the conditions of  $\rho(t)$  and  $v(k, t)$ .

## 5. Influence of PGS on Pedestrians

Guiding information intensity shows the impact level that guiding signs have on the pedestrian in the specific area. The influence level has a connection with the guiding distance, the guiding information, and the guiding destination. As it is

shown in Figure 4, area  $\delta_1$  refers to the area where guiding signs have effects on optical field. It means that, within the scope of  $\delta_1$ , people could see the guiding sign and probably approach their destination. On the other hand,  $\delta_2$  shows the guiding effect area. It means that, in this area, people confirm whether the signs could guide.

$R(k, s, d)$  represents the degree of guiding impact of the guiding sign on the  $k$  class pedestrians in the direction of  $d$  at the distance of  $s$ :

$$R(k, s, d) = \begin{cases} 0, & s > \delta_1, \\ \lambda_{v1}(s) \cdot d(k) \cdot \tan \theta_v(d), & \delta_2 \leq s \leq \delta_1, \\ \lambda_{v2}(s) \cdot d(k) \cdot \tan \theta_l(d), & s \leq \delta_2, \end{cases} \quad (4)$$

where  $\lambda_{v1}(s)$  denotes the guiding coefficient in the visual impact area ( $\delta_2 \leq s \leq \delta_1$ ) of the guiding sign.  $\lambda_{v2}(s)$  denotes the guiding coefficient in the visual impact area ( $s \leq \delta_2$ ) of the guiding sign, which is a function of the distance  $s$  and  $d(k)$  denotes the destination direction of the  $k$  class pedestrians.  $\theta_v(d)$  denotes the intersection angle of the destination direction and the connection line between the pedestrian and the guiding sign.  $\theta_l(d)$  denotes the intersection angle of the guiding direction and the destination direction of the pedestrian. In the range of  $\delta_2 \leq s \leq \delta_1$ , the main role of the guiding signs is to attract pedestrians to further confirm the position of the guiding signs. In the range of  $s \leq \delta_2$ , the main role of the guiding signs is to show the path of the pedestrian's present position and the destination based on the direction showed on the guiding signs.

## 6. Solution to Model Simulation

All the models above will be used in this part to describe the actual procedure in airport.

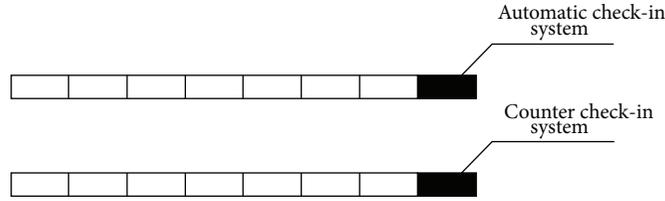


FIGURE 5: Cellular automation model.

### 6.1. Checking-In Procedures Module

**6.1.1. Choice of Cellular.** CA model was introduced here to describe the checking-in procedures. Every line in front of a checking-in will be represented by a line of cellular. The beginner of the list is the checking-in counter. Here, the counter includes the automatic machine and the checking-in counter, which are shown in Figure 5.

**6.1.2. Construction of Cellular.** As is shown in Figure 5, every cellular in the line represents a passenger walk-in. The front of the line is the checking-in counter. The automatic checking-in will cost a random amount of time, the average of which will be 30 s. The check-in counter will be used up a random amount of times too, the average of which will be 80 s.

**6.1.3. The Running of Cellular Automation.** The initial time for the simulation is set as  $t_0$ . For each cellular, simulation length is 0.7 m, the length of the queuing region is  $100 \times 0.7$  m, and the walking speed of the passenger is set to be distance of one cellular per second, that is, 0.7 m/s according to the given document. At the front of the queue, cellular is released every 30 s or 80 s to simulate the process of check-in formalities.

(1) *Phase 1: Cellular Initialization.* Initialize empty one-dimensional cellular. One-dimensional cellular queue produces the number of the tourists needing to register corresponding to each simulation step based on the module of the side lane, parking lot, public transportation, and the passenger capacity distributed over time which needs to register output by the module of side waiting lounge.

(2) *Phase 2: The Running of Cellular Automation*

(a) *Running Principle.* Cellular Automation will be used to simulate the situation that passengers check in system. Each simulation moment is one step size; the cellular automata run at a constant speed  $v = 1$ , as showed in Figure 6.

(b) *Rules for Passengers Changing a Line.* When a passenger finds extra lines are shorter than the one that he is waiting in, the passenger will get into the line he prefers, as is shown in Figure 7.

(c) *Update the Queue.* For cellular at the moment  $t_p$ , its position state in the queue for the next second is

$$x(i, t_p + 1) = x(i, t_p) + v(i, t_p + 1). \quad (5)$$

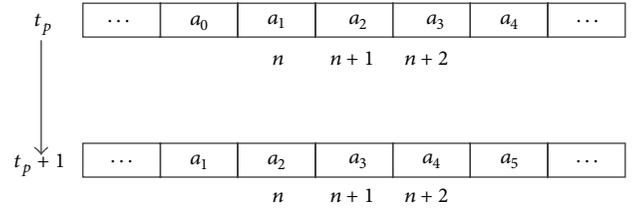


FIGURE 6: One-dimensional cellular movement rules of queues.

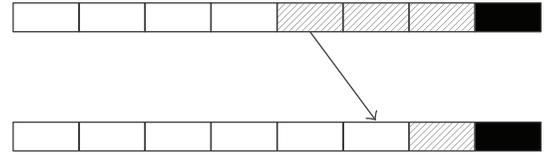


FIGURE 7: Schematic of passengers changing a line.

### 6.2. Security Checking Module

**6.2.1. Basic Procedures.** Security check is an important part of the simulation of pedestrians inside the terminal. The guiding sign of security check has a significant impact on pedestrians' behaviors. Process is as shown in Figure 8.

**6.2.2. Choice of Cellular.** We show every waiting line in security checking area as a line of cellular. In the simulation, security procedures are separated into two types: single queue single serve (SQSS) and single queue double serve (SQDS) (as shown in Figure 9). Correspondingly, the first one or two cellars represent the passengers taking security checking.

**6.2.3. Consistence of a Cell.** In the simulation process, each cell represents an area of  $0.7 \text{ m} \times 0.7 \text{ m}$  in the security check area. As shown in Figure 10, each passenger waiting for the security check or being checked occupies a cell, corresponding to the two situations of single queue single check and single queue double check. The first one or two cells represent passengers taking security checking. Passenger walking speed is set to 0.7 m/s in accordance with the applicable documents, which denotes that passengers move forward the distance of a cell's length per second.

**6.2.4. The Running of Cellular Automation.** In this simulation, the maximum queue length is set at  $50 \times 0.7 \text{ m}$ , the

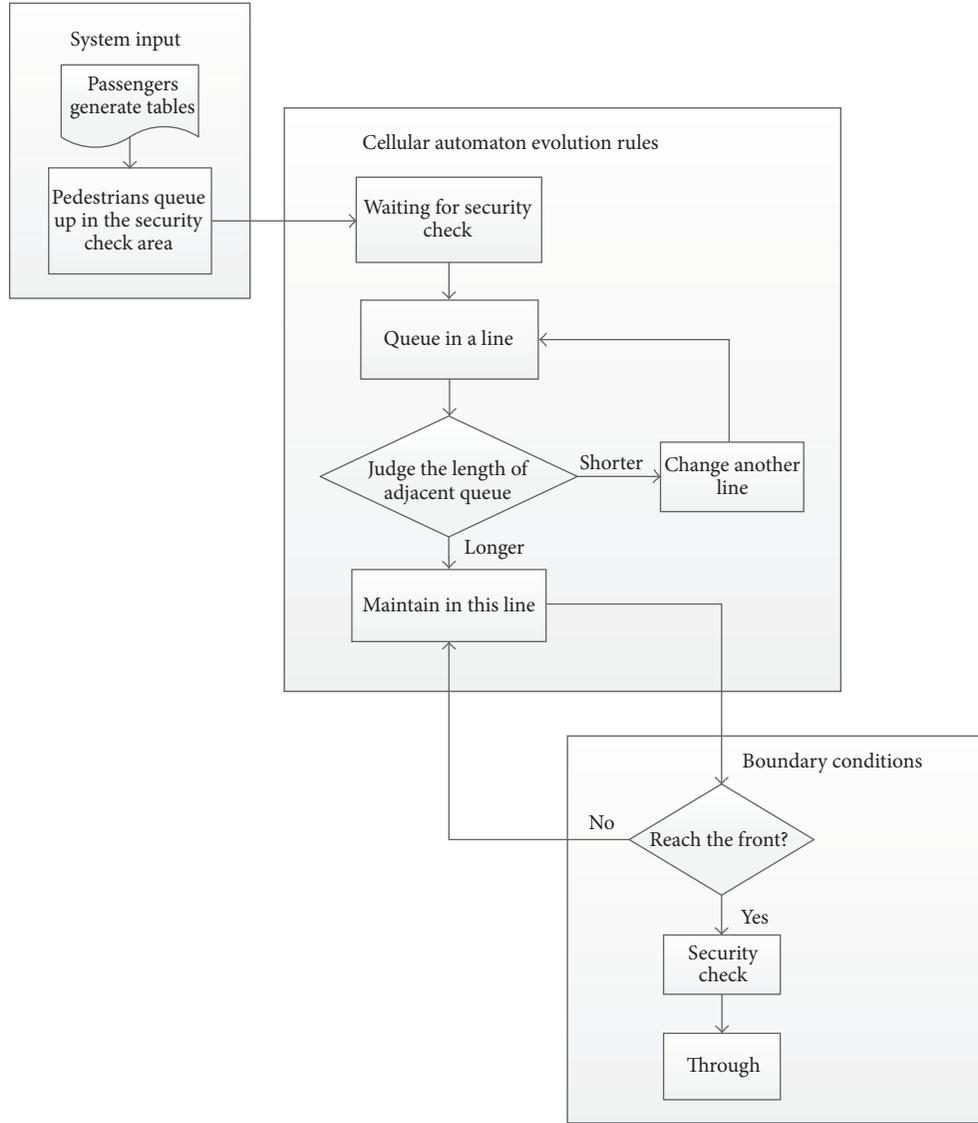


FIGURE 8: Process of passengers' security check system.

simulation step is 1 second, and the total simulation time duration is 4 hours.

(1) *Phase 1: Cellular Initialization.* Maximum queue length and the gapping security check counter number  $n$  form a  $50 \times n$  cellular space, representing a complete security queuing area; at the initial time the cells in the entire cellular space are not occupied.

Thus, the  $300 \times n$  matrix is generated according to the total number of passengers being checked every five minutes (no airline distinction), that is, the normal distribution matrix of the passengers coming into the security check area within five minutes. The next 300 simulation steps (5-minute simulation time) read the matrix line by line. All the passengers coming into the security check area within five minutes enter the security queuing area randomly during the simulation time.

(2) *Phase 2: The Running of Cellular Automation*

(a) *Basic Evolution Principle.* The basic cellular evolution rule is, if the adjacent cellular in front of the cell is empty, then the cell moves forward a position. This evolution rules can be applied to the following situations: first, passengers entering the security queuing area but not yet coming into a queue; second, passengers waiting in a queue and the neighbor queues are not shorter than the one of the passengers. The premise of evolution in accordance with such rules which does not allow jumping phenomenon.

(b) *Rules of Passenger Changing a Line.* If there is one team shorter that the other in two neighboring lines and the corresponding cellular to the passenger is at the same time empty, then the passenger in the corresponding cellular will change from the current queue to the shorter one; if both sides of the two adjacent queues are shorter than the current queue

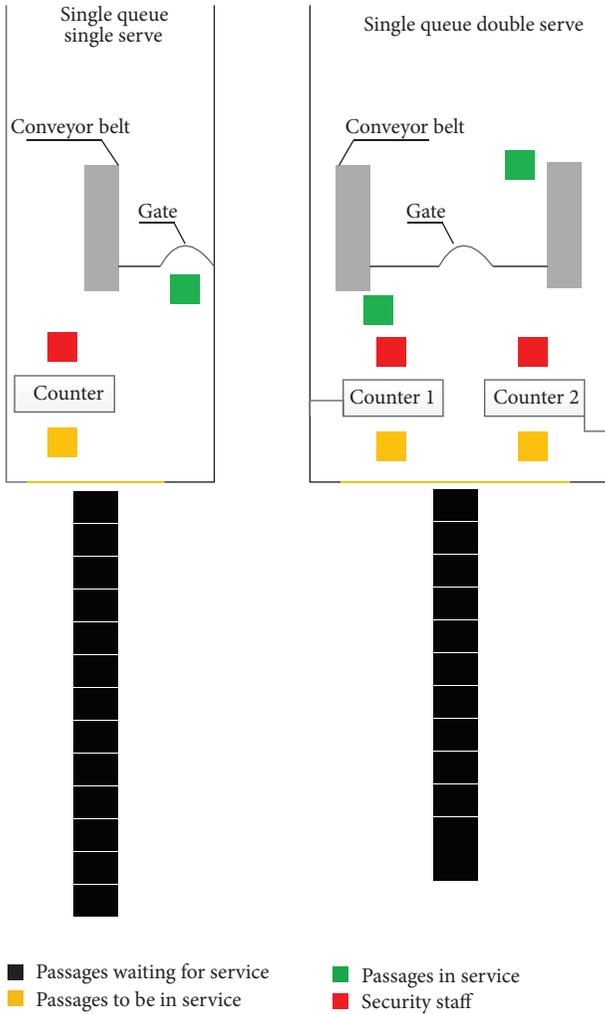


FIGURE 9: Chart of single queue single serve (SQSS) and single queue double serve (SQDS) condition.

the passenger is in and the corresponding cellulars in both queues are empty, the passenger will choose to change into one shorter queue randomly. Details are shown in Figure 11.

(c) *Boundary Condition.* The hypothesis is that all passengers pass through the security. After the passenger at the head of the queue passes the security check, it denotes the time that the corresponding cellular occupies this position reaches the time of the security check, so the passenger will leave the system through the boundary.

### 6.3. Airport Lounge Module

6.3.1. *Inner Airport Lounge Module Cellular Automation Model.* Inner airport lounge is regarded as a single cell, several security check is gathered as an entrance—a one-dimensional cellular automata I connected to the cellular of an airport lounge. Boarding channel exits are summarized as the exit and seen as one-dimensional cellular connected with the cellular in the waiting hall. The result is the construction

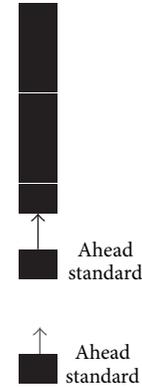


FIGURE 10: Basic principle for waiting cellular in the security area.

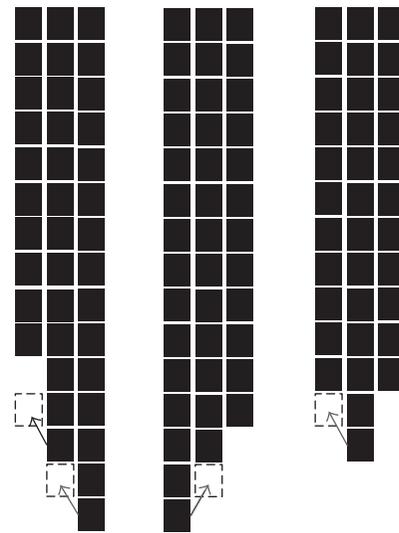


FIGURE 11: Chart showing passenger changing a line.

of one cellular automaton model which has a single entrance and a single exit and is connected to a single one-dimensional cellular.

6.3.2. *Consistence of Cellular.* As showed in Figure 11, the one-dimensional cellulars at the entrances are all composed of numerous cellulars. Each cellular represents the number of passengers going through all the security checkpoints at a simulation step  $t_1$ . The value of the cellular is the number of the group of people.

The one-dimensional cellular at the entrances initialized with the unit of the simulation step, in accordance with the passengers output by the security check module for the distribution of time. The one-dimensional cellular with the exit starts boarding half an hour earlier according to the flight schedule departure time. The one-dimensional cellular is initialized according to the average delay of 4–6 seconds per passenger while getting through the export.

6.3.3. *The Running of Cellular Automation.* The number of people in the waiting hall is  $room\_origin$ . The simulation

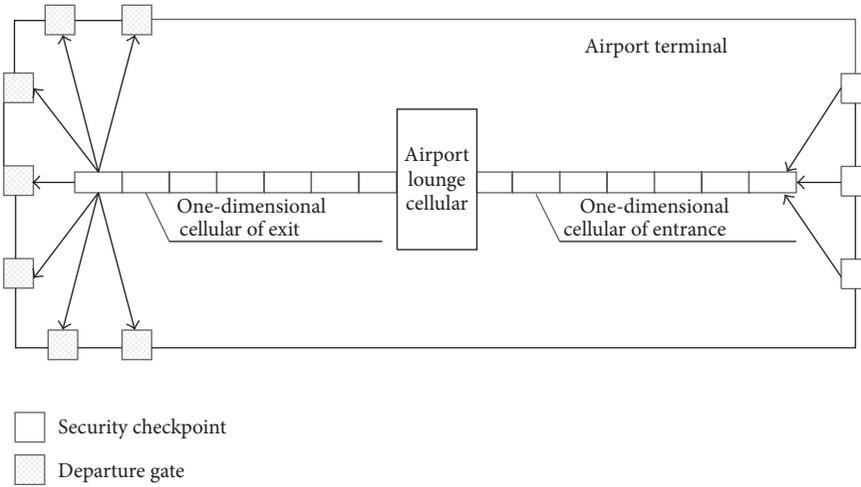


FIGURE 12: Airport lounge cellular automation chart.

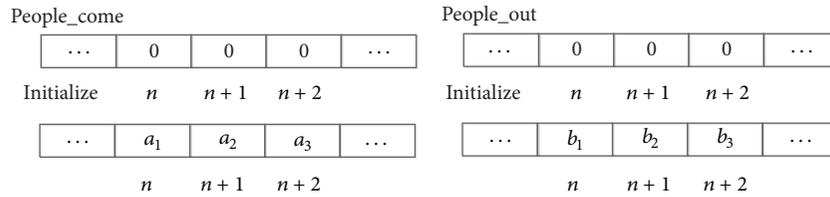


FIGURE 13: Initialization chart of one-dimensional cellular at the entrance and export.

time is  $t_n$  step. The one-dimensional cellular *people\_come* at the entrance is composed of  $1 \times t_n$  cellulars. The cellular *people\_come* ( $n$ ) denotes the number of people passing the security check points into the waiting hall at the  $n$ th simulation time. The one-dimensional cellular *people\_out* of the export is composed of  $1 \times t_n$  cellular and the cellular *people\_out* denotes ( $n$ ) the number of people leaving the waiting hall and starting boarding at the  $n$ th simulation time.

(1) *Phase 1: Cellular Initialization.* After the initialization of the empty one-dimensional cellular, according to the output number of passengers distributed over time by security check module, the one-dimensional cellular *people\_come* obtained the number of arriving passengers  $a_1, a_2, a_3, \dots$ , corresponding to each simulation step *people\_come* ( $n$ ) *people\_come* ( $n + 1$ ), *people\_come* ( $n + 2$ )  $\dots$ , as Figure 12 shows. According to the flight schedule departure time, namely, starting boarding half an hour earlier rule and the average delay of each passenger passing the export, we initialize the export one-dimensional cellular, getting the number of people  $b_1, b_2, b_3 \dots$  which leave the waiting hall and start boarding corresponding to each simulation step *people\_out* ( $n$ ), *people\_out* ( $n + 1$ ), *people\_out* ( $n + 2$ ),  $\dots$ , as Figure 13 shows.

(2) *Phase 2: The Running of Cellular Automation*

(a) Operating rules: the cellular automaton simulates the situation of passengers entering or leaving the waiting hall. Each simulation time is one step. Therefore, the cellular automaton runs at a constant speed  $v = 1$ . So cellular changes from the simulation time  $t_p$  to  $t_{p+1}$  as shown in Figure 14.

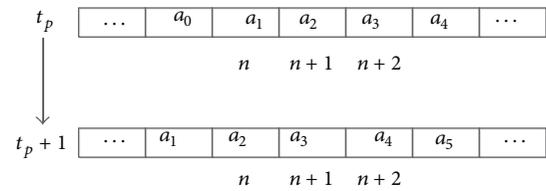


FIGURE 14: Chart of one-dimensional cellular movement at the entrance and export.

(b) The update of the number of people in the airport lounge: at the initial time  $t_0$  the number of people in the airport lounge is *room\_origin*. The cellular to the export is *people\_come*, *people\_out*. So at the initial moment, the number of people in the airport lounge *room\_people* can be described as

$$\begin{aligned} \text{room\_people}(t_0 + 1) &= \text{room\_origin} + \text{people\_come}(1) \\ &\quad - \text{people\_out}(1). \end{aligned} \tag{6}$$

Till  $t_{p+1}$ , the number of people in the waiting hall *room\_people* can be expressed as

$$\begin{aligned} \text{room\_people}(t_0 + t_p + 1) &= \text{room\_people}(t_0 + t_p) + \text{people\_come}(t_p) \\ &\quad - \text{people\_out}(t_p). \end{aligned} \tag{7}$$

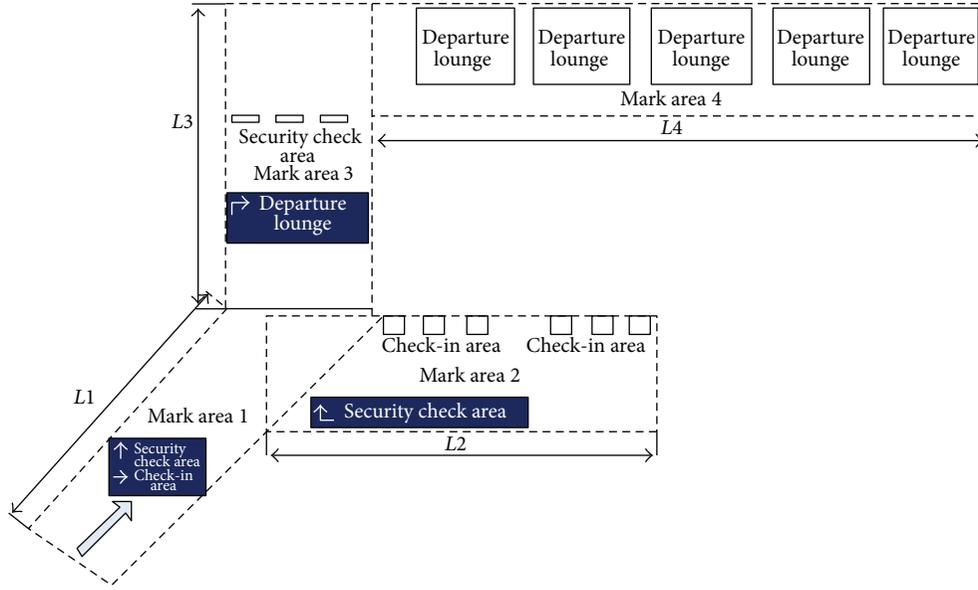


FIGURE 15: Chart of transportation facilities inside the terminal.

TABLE 1: Experiment 1: changing guiding signs' height.

Order	Height (meters)	Rate of flow (people/hour)	Intensity (number/50 meters)
1	2.3	1000	3
		1500	
		2000	
		2500	
2	2.6	1000	3
		1500	
		2000	
		2500	
3	2.9	1000	3
		1500	
		2000	
		2500	
4	3.2	1000	3
		1500	
		2000	
		2500	

TABLE 2: Experiment 2: changing guiding signs' intensity.

Order	Intensity (/50 meters)	Rate of flow (people/hour)	Height (meters)
1	2	1000	2.6
		1500	
		2000	
		2500	
2	3	1000	2.6
		1500	
		2000	
		2500	
3	4	1000	2.6
		1500	
		2000	
		2500	
4	5	1000	2.6
		1500	
		2000	
		2500	

## 7. Model Application

7.1. *Simulation Environment.* A simulation scene consists of passage, checking-in area, security check, and the waiting area. The normal procedure is passage—checking-in counter—security check—waiting—boarding, several of which need to be guided by signs. If signs were not set up correctly, pedestrians will be held up and all procedures will take longer. This will lay pressure on terminal's bearing. In simulation scene, lengths of  $L1$ ,  $L2$ ,  $L3$ , and  $L4$  are 40 m, 50 m, 50 m, and 100 m.

Transportation facilities chart inside the emulated terminal are shown in Figure 15. PGS adopt suspension type. The number and the height will be set up in accordance with requirements of different experiment.

7.2. *Simulation Experiment Design.* Based on the analysis, the intensity and height of transportation signs have an influence on pedestrians' routing selection, which is also related to the number of pedestrians. This thesis paper separates the abovementioned variations into two parts: guiding signs number optimization and guiding signs height optimization. The simulation experiment is shown in Tables 1 and 2.

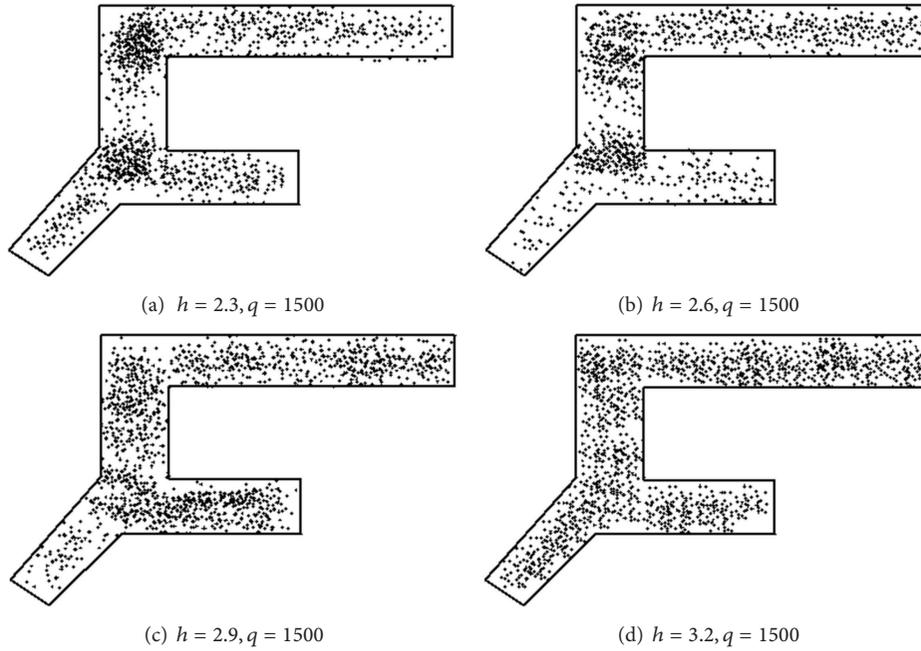


FIGURE 16: Inner space status of signs at different heights ( $q = 1500$ ).

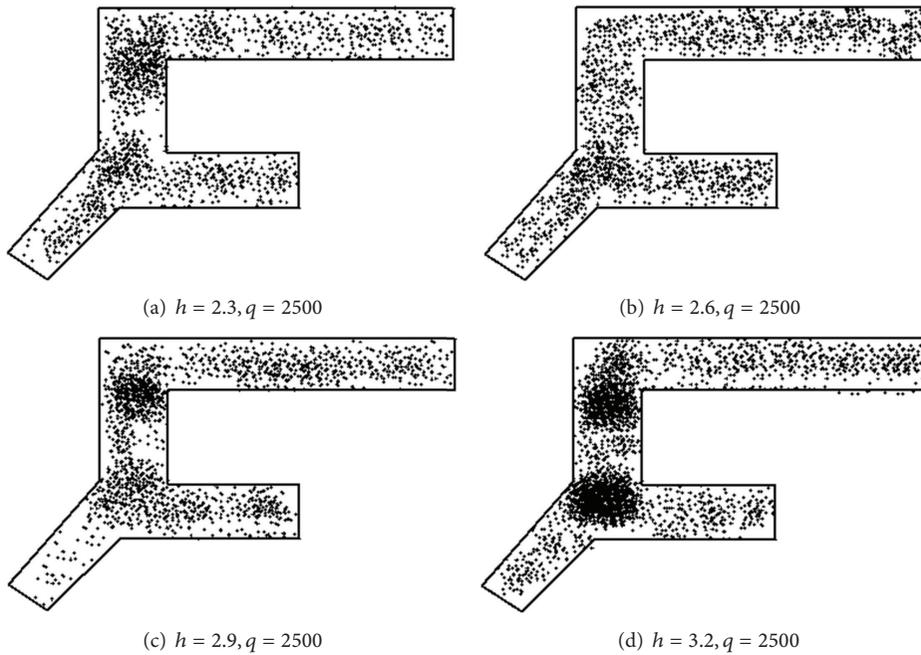


FIGURE 17: Inner space status of signs at different intensity levels ( $q = 2500$ ).

### 7.3. Result of Simulation Experiment

7.3.1. *Experiment 1 Result of Simulation.* Based on simulation experiment number 1, terminal space variation in three minutes is shown as Figures 16 and 17 ( $h =$  height;  $q =$  rate of pedestrian flow).

Under the condition of experiment number 1, the contrast Chart between the delay for signs of identification and

the travel length to the airport lounge is shown as in Figure 18.

7.3.2. *Simulation Result from Experiment Number 2.* Conducted under the conditions of experiment number 2, terminal space variation for 3 minutes is shown as Figure 19 ( $\rho =$  intensity,  $q =$  rate of pedestrian flow).

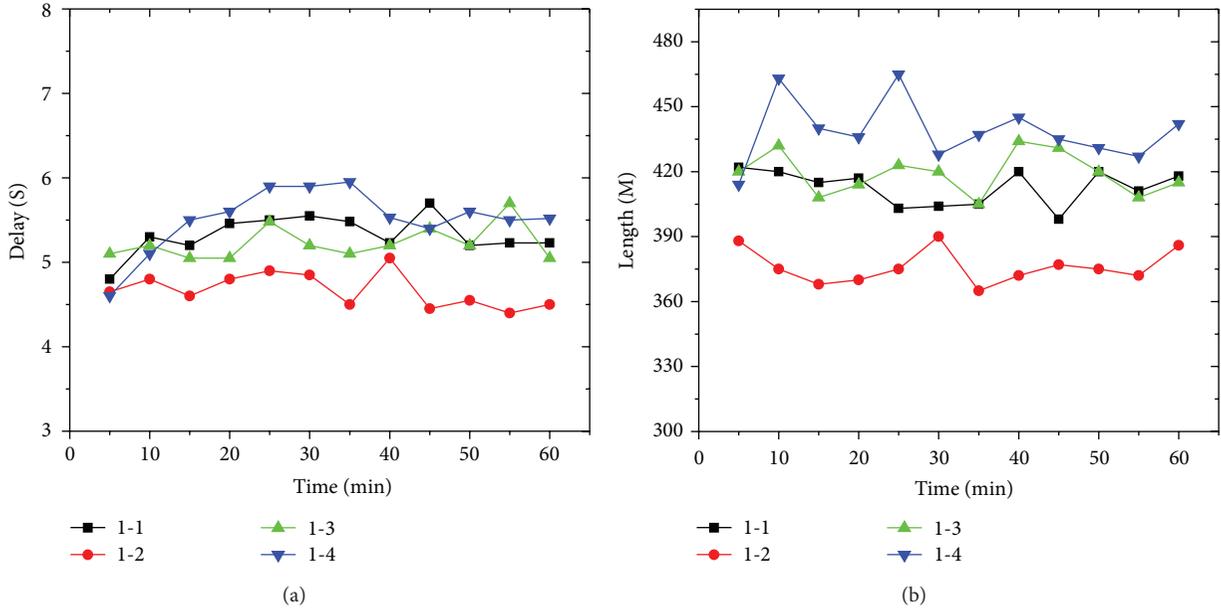


FIGURE 18: Contrast of delay and travel length (number 1).

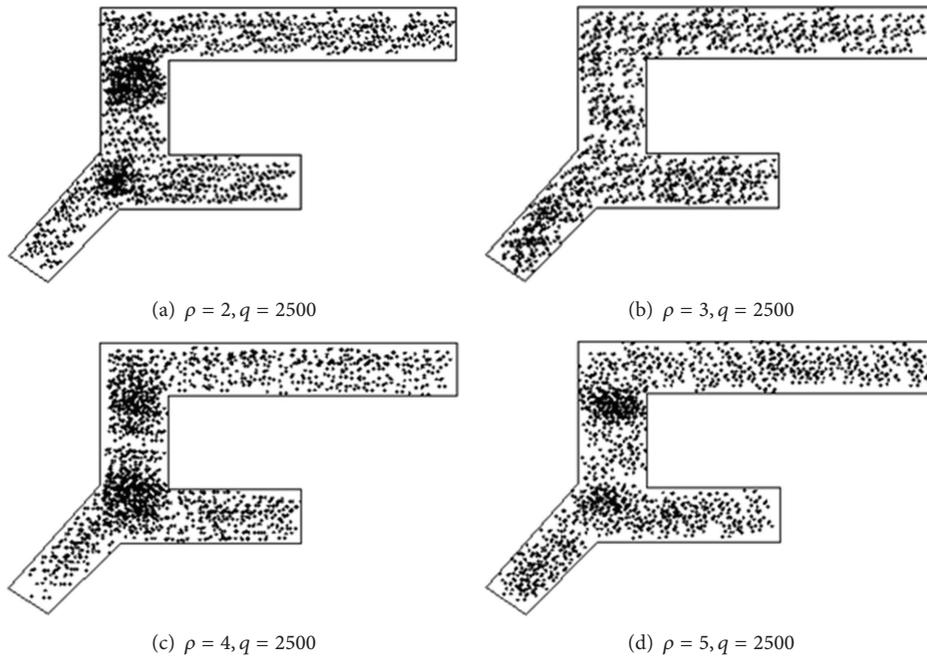


FIGURE 19: Inner space status sat different intensities (number 2).

Under the condition of experiment number 2, the contrast Chart between the delay for signs of identification and the passage length to the airport lounge is shown as in Figure 20.

7.4. Analysis of the Result of the Simulation Experiment. (1) From the simulated inner space status chart resulting from experiment number 1, when the rate is 1500 p/h, the height of guiding signs does not have a substantial influence

on pedestrians. No obvious grouping is observed for routing decision-making area under different heights. When the rate of pedestrian flow is 2500 p/h, different heights have effects on routing decisions. Grouping level is different at different locations under different conditions. Either too high or too low will cause people to stay with the routing selecting area. From simulation experiment number 1, it is found that a difference of routing selection comes from the guiding sign heights. Under this condition of experiment number 1, the recommendation to the guiding sign heights is 2.6 m.

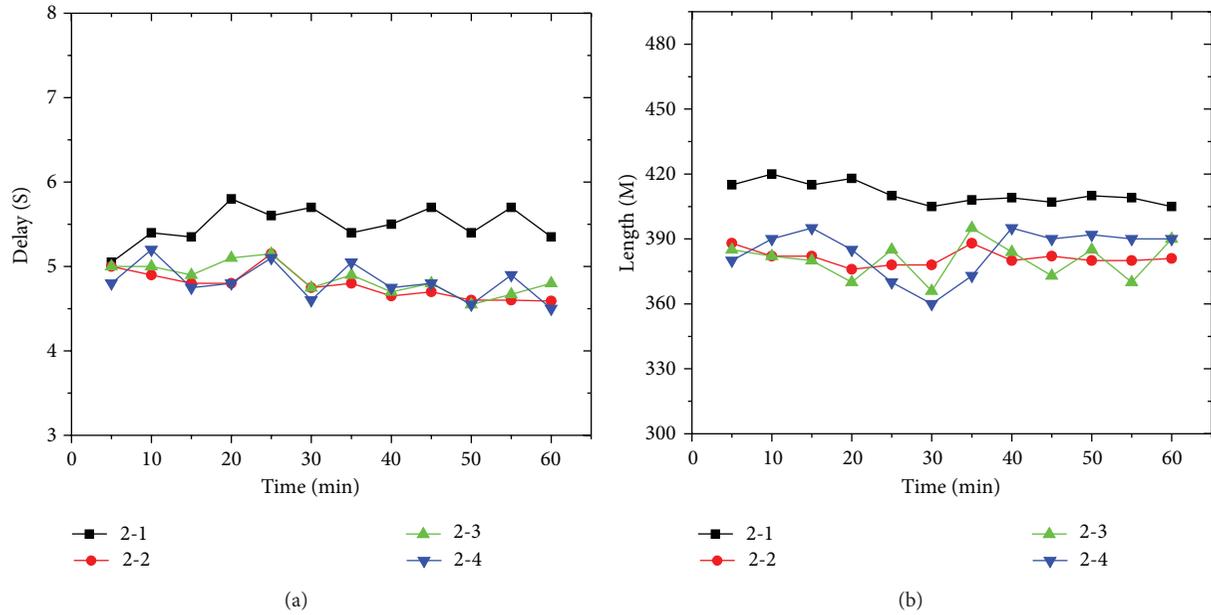


FIGURE 20: Contrast of delay and travel length (number 2).

(2) From the simulated inner space status chart resulting from experiment number 2, when the rate is 2500 p/h, discrete intensity levels have effects on routing selection and differences are found in grouping level at varied positions under different conditions. When the intervals are too large, pedestrians will spend more time looking for signs and also will possibly make a mistake choosing direction due to lack of sign. And furthermore, local space of pedestrians will happen. When the intervals are too small, pedestrians might have problems or additional burdens in identifying signs. From the above, the recommended guiding sign intensity is 3 signs per 50 m.

## 8. Conclusions

In this paper, an optimization method of PGS is formulated. Pedestrian traffic simulation procedures within the AT are outlined. Firstly, according to features of the PGS' appearances, the boundary judge and feature perception models are built in order to simulate the properties of the pedestrian. Then, a self-organization model is formulated to determine herd behavior features of the pedestrians. Thirdly, the CA model is used to simulate the crowd situation of the pedestrian within the AT. The result demonstrates that the method in this paper could analyze the influence of the height of PGS on routing decision-making of the pedestrian in AT. A reasonable height for the PGS is recommended for using in the AT according to the numerical simulation result. Meanwhile, the model could describe the influences of PGS's density on the pedestrian routing decisions. We hope that the models presented in this paper could provide additional guidance for optimizing the interior traffic facilities within the AT.

Future research is directed towards developing methods to solve the guiding signs optimization problem efficiently

and widely. Moreover, practical applications of the approach will be taken into account, such as the route choice of guiding by acoustic identification and other guiding facilities.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

The research is supported by the National Basic Research Program of China (no. 2012CB725403).

## References

- [1] W. Kim, Y. Park, and B. J. Kim, "Estimating hourly variations in passenger volume at airports using dwelling time distributions," *Journal of Air Transport Management*, vol. 10, no. 6, pp. 395–400, 2004.
- [2] G. F. Newell, *Application of Queuing Theory*, Chapman and Lounge, London, UK, 1971.
- [3] W. J. Dunlay Jr. and C. H. Park, "Tandem-queue algorithm for airport user flows," *Transportation Engineering Journal of ASCE*, vol. 104, no. 2, pp. 131–149, 1978.
- [4] S. Solak, J.-P. B. Clarke, and E. L. Johnson, "Airport terminal capacity planning," *Transportation Research B*, vol. 43, no. 6, pp. 659–676, 2009.
- [5] I. E. Manataki and K. G. Zografos, "A generic system dynamics based tool for airport terminal performance analysis," *Transportation Research C*, vol. 17, no. 4, pp. 428–443, 2009.
- [6] A. S. Kiran, T. Cetinkaya, and S. Og, "Simulation modeling and analysis of a new international terminal," in *Proceedings of the Winter Simulation Conference*, pp. 1168–1172, Society

- for Computer Simulation International, Orlando, Fla, USA, December 2000.
- [7] M. Schultz and H. Fricke, "Managing passenger handling at airport terminals," in *Proceedings of the 9th USA/Europe Air Traffic Management Research and Development Seminar*, Berlin, Germany, 2011.
  - [8] P. K. Chawdhry, "Risk modeling and simulation of airport passenger departures process," in *Proceedings of the Winter Simulation Conference (WSC '09)*, pp. 2820–2831, December 2009.
  - [9] I. Akgun, A. Kandakoglu, and A. F. Ozok, "Fuzzy integrated vulnerability assessment model for critical facilities in combating the terrorism," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3561–3573, 2010.
  - [10] C.-C. Chou, "A model for the evaluation of airport service quality," *Proceedings of the Institution of Civil Engineers: Transport*, vol. 162, no. 4, pp. 207–213, 2009.
  - [11] A. G. de Barros, A. K. Somasundaraswaran, and S. C. Wirasinghe, "Evaluation of level of service for transfer passengers at airports," *Journal of Air Transport Management*, vol. 13, no. 5, pp. 293–298, 2007.
  - [12] Y. Park, "A methodology for establishing operational standards of airport passenger terminals," *Journal of Air Transport Management*, vol. 5, no. 2, pp. 73–80, 1999.
  - [13] Y. H. Park, *An evaluation methodology for the level of service at the airport landside system [Ph.D. thesis]*, Loughborough University of Technology, Loughborough, UK, 1994.
  - [14] J.-R. Yen, "A new approach to measure the level of service of procedures in the airport landside," *Transportation Planning Journal*, vol. 24, no. 3, pp. 323–336, 1995.
  - [15] Guidelines for Transit Facility Signing and Graphics TCRP Report 12, 1996.
  - [16] A. Uebele, *Signage Systems & Information Graphics: A Professional Sourcebook*, Thames & Hudson, 2007.
  - [17] M. L. Tam, "An optimization model for wayfinding problems in terminal building," *Journal of Air Transport Management*, vol. 17, no. 2, pp. 74–79, 2011.

*Research Article*

# **A Bilevel Programming Model to Optimize Train Operation Based on Satisfaction for an Intercity Rail Line**

**Zhipeng Huang and Huimin Niu**

*School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China*

Correspondence should be addressed to Huimin Niu; hmniu@mail.lzjtu.cn

Received 12 December 2013; Revised 5 January 2014; Accepted 14 January 2014; Published 20 February 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Z. Huang and H. Niu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The passenger travel demands for intercity rail lines fluctuate obviously during different time periods, which makes the rail departments unable to establish an even train operation scheme. This paper considers an optimization problem for train operations which respond to passenger travel demands of different periods in intercity rail lines. A satisfactory function of passenger travelling is proposed by means of analyzing the passengers' travel choice behavior and correlative influencing factors. On this basis, the paper formulates a bilevel programming model which maximizes interests of railway enterprises and travelling satisfaction of each passenger. The trains operation in different periods can be optimized through upper layer planning of the model, while considering the passenger flow distribution problem based on the Wardrop user equilibrium principle in the lower layer planning. Then, a genetic algorithm is designed according to model features for solving the upper layering. The Frank-Wolfe algorithm is used for solving the lower layer planning. Finally, a numerical example is provided to demonstrate the application of the method proposed in this paper.

## **1. Introduction**

On intercity rail lines, the train dispatching density is quite high and passengers normally choose their travel time according to their own travel habits. The passenger travel demand varies at different time periods, and the distribution of passenger flow during the day has obvious peaks and troughs. The railway department will fully consider the characteristic of passengers' travel time when setting the train operation scheme of intercity rail lines. If the evened train dispatching mode is adopted for the whole day, it will cause the dissatisfaction of the demands in some peak periods and excess supply in some trough periods. Therefore, to divide the service time into several periods and adopt the evenly dispatching mode based on period will make a better balance between passengers' travel demand and benefits of railway transportation. In addition, passengers will also consider traveling in high-satisfactory time periods according to the train operation scheme. Therefore, to develop a reasonable operation scheme can effectively eliminate the shortage issue of dispatching capacity on stations during peak hours, also

leading passengers to travel at different time periods without reducing their travel satisfactions.

Domestic and overseas scholars have done a great amount of researches on the passenger train operation scheme of railway transportation. Considering that the traveling characteristic of intercity passenger flow is different from that of ordinary railway, some scholars have analyzed and researched passenger flow characteristics of intercity railway. Through massive investigation data, Mandel et al. have done a systematic research on the high-speed railway market and analyzed the fact that the characteristics of passengers' demand are similar to those of city bus passenger demands on high-speed railway in Germany [1]. Zhao has systematically analyzed the fact that the traveling characteristics of intercity railway passenger flow are remarkably different from those of the urban residents in the aspects of traveling objective, time, intensity, distance, cost, and distribution [2]. Through numerous researches by domestic and overseas scholars, it is found that the passenger flow of intercity railway has a higher variability, on the elastic demand of which some scholars have done researches. Cascetta and Coppola have

established an optimized model of train operation timetable under the elastic demand, aiming to optimize the ticket pricing structure of high-speed railway [3]. Shi et al. have analyzed the affecting factors related to the passenger train operation scheme and constructed a function of elastic passengers' transport demand [4]. Some scholars believe that the reasonable distribution of passenger flow is the premise of optimizing the passenger train operation scheme. Peng and Zhu have described the interactive relationship between transport demand and the operation scheme, established a dynamic assignment model of passenger flow which reflects the transition process for different types of passenger flow, and constructed multiobjective planning model which gives a comprehensive consideration to maximum benefit of railway department, minimum total cost of passenger, and satisfaction of passenger flow demand to the greatest extent [5]. Some scholars have given enough thought to the impact of congestion on passengers' travel when optimizing the train operation scheme. Niu and Zhou have studied the optimizing issue of train timetable for city rail transportation under time-varying and congestion condition. Using time-varying based passenger flow data collected from the automatic ticket inspection system of city rail transportation, they established the nonlinear integer planning model for dispatching trains on rail under the given condition by the number of motor train units [6]. There are many issues to be taken into account in order to develop a reasonable passenger train operation scheme, and the various factors often have association relationship with each other. Therefore, some scholars have tried to establish bilevel planning model to give mathematical description to this issue and obtained certain achievements. He et al. have analyzed the balancing condition of passenger flow for different types of trains and established the generalized cost function for different types of trains, based on which they have constructed multiobjective bilevel planning model for developing the dedicated passenger train operation scheme. This model takes the maximum benefit of train operation and maximum convenient degree of passenger as target and also considers the assignment issue of different types of trains based on minimum travel cost to passengers [7]. Shi et al. have analyzed the relevant cost and optimizing objective of the passenger train operation scheme, evenly considered the enterprise profit and the passengers' demand, combined the train operation scheme and passenger transference scheme, and established the optimized bilevel planning model of the passenger train operation scheme which takes distribution of passenger flow in railway transference network as the lower layer planning [8]. Considering that the intercity railway train operation scheme may affect the travel convenience of passenger flow and passenger's satisfactory degree, He et al. have studied the convenience for passengers in different time periods based on market investigation data [7]. Chang et al. developed a multiobjective programming model for the optimal allocation of passenger train service on an intercity high-speed rail line without branches. Minimizing the operator's total operating cost and minimizing the passenger's total travel time loss are the two planning objectives of the model [9]. Huang and Niu have

analyzed the relevant factors that affected the traveling satisfaction of passengers and established an optimized model of train operation scheme aiming at getting maximum traveling satisfaction of passengers [10]. However, the above research did not consider the effect of passenger flow distribution for passengers' satisfaction and the feedback relationship between train operations scheme and passengers' demand. Therefore, this paper will construct a bilevel programming model to optimize train operations based on passengers' satisfaction.

The remainder of this paper is organized as follows: Section 2 presents and analyzes the researching issues; Section 3 establishes bilevel planning model of intercity railway train operation scheme; Section 4 gets solution by adopting genetic algorithm based on model features; Section 5 verifies the model through examples; in the end, the content of study in this paper and the further issues to be researched are summarized.

## 2. Problem Statement

*2.1. Division of Passenger Travelling Time Period.* On intercity rail lines, different passengers will choose different travelling time periods based on their own demands. For example, some passengers will make decisions on travelling time in accordance with their habits; some will do so according to the congestion level of urban traffic, while some will do it as requested by their working hours and so on. However, through the research on the travelling time period of intercity railway passengers, it shows that passenger flow is not distributed equally in different time periods but has obvious peaks and troughs throughout the day.

According to the above analysis, the passengers' travel demand is uneven in different time periods. Therefore, the passenger travel demand in different time periods needs to be taken into account when developing the intercity railway train operation scheme. In order to make an easy study, this paper divides the service time of intercity rail lines into  $m$  travel periods by hour; that is, one hour represents one travel period.  $P$  serves as the set of travelling time periods and  $P \in \{p \mid 1, 2, \dots, m\}$ .

*2.2. Generalized Passenger Travel Cost.* Passengers must pay the cost of ticket and time when travelling. Thereof, time cost consists of two parts: waiting time cost and travelling time cost. Waiting time cost means the average waiting time after the passenger arrives at the station; travelling time cost is the time period that passenger has to spend in order to complete the spatial displacement, which remains unchanged when the train speed is constant. Therefore, only the waiting time cost will be taken into consideration in this paper. Meanwhile, as congestion causes inconvenience for passenger's travelling, congestion cost is used to quantize the passenger's travelling convenient degree in this paper. The generalized passenger travel cost defined in this paper includes three parts: ticket cost, waiting time cost, and congestion cost, as shown in

$$V_p = C + \lambda \cdot t_p + k_p. \quad (1)$$

Thereof,  $V_p$  is the generalized travel cost of each passenger in the time period  $p$ ;  $C$  is the ticket cost;  $\lambda$  is the time value of passenger;  $t_p$  is the average waiting time of each passenger in the travelling time period  $p$  (the variation of the value is related to the train dispatching intervals, as shown in formula (2));  $k_p$  is the congestion cost for the passengers in the travelling time period  $p$  (the value is on the basis of ticket cost, as shown in formula (3)). We have

$$t_p = \frac{1}{2} \cdot I_p, \quad (2)$$

where  $I_p$  is the train dispatching intervals in the travelling time period  $p$ , and

$$k_p = C \cdot \exp\left(-\frac{N}{Q_p}\right), \quad (3)$$

where  $Q_p$  is the passenger flow volume which is assigned in the travelling time period  $p$ ;  $N$  is the service capacity of the station;  $\exp(i)$  is the natural exponential function. The greater the  $N/Q_p$ , the lower the passenger congestion (the value  $\exp(-N/Q_p)$  at this point gets closer to zero); the smaller the  $N/Q_p$ , the higher the level of passenger congestion (the value  $\exp(-N/Q_p)$  at this point gets closer to 1).

**2.3. Passenger's Travelling Satisfaction.** The passenger's travelling satisfaction defined in this paper means the satisfaction level of expected travelling time and the generalized travel cost for all passengers in the same time period. The factors affecting passenger's travelling satisfaction include two parts as follows.

(1) *Average Satisfaction Level of Travelling Time Period.* As previously mentioned, different passengers will choose different travelling time periods according to their own demands. The key factor affecting the passenger's satisfaction is whether enough transportation capacity could be provided by the railway department according to the passenger's demand of travelling time period. The average passenger satisfaction  $A_p$  of the travelling time period  $p$  is

$$A_p = \begin{cases} 1 & Q_p \geq M_p, \\ \frac{Q_p}{M_p} & Q_p < M_p, \end{cases} \quad (4)$$

where  $M_p$  is the passenger demand of the travelling time period  $p$ , which can be obtained through market survey; when  $Q_p \geq M_p$ , it indicates all passengers expecting to travel at the  $p$  time period, whose demands on travelling time can be satisfied. When  $Q_p < M_p$ , it indicates that some passengers have to choose other time periods, and therefore the average satisfaction of the passengers will decrease. The value range of  $Q_p/M_p$  is  $[0, 1)$ . The smaller the  $Q_p$  is, the lower the passenger's average satisfaction gets.

(2) *Passenger Satisfaction of Generalized Travel Cost.* Passengers always expect the minimum travel cost. The lower the generalized cost of passenger's travelling is, the higher its

satisfaction gets. The passenger's travelling satisfaction  $B_p$  of generalized cost in the travelling time period  $p$  is

$$B_p = 1 - \frac{V_p - C}{V_p}. \quad (5)$$

As mentioned above, the ticket cost  $C$  that passenger has to pay when travelling is a set value, while the generalized cost  $V_p$  paid by passenger at the travelling time period  $p$  is a variable value.  $V_p - C$  shows the extra cost paid by passenger in addition to the ticket cost. The more this part costs, the lower the passenger's satisfaction level gets.

Through the above analysis, the passenger's travelling satisfaction presents a dynamic balance. When passenger can travel according to his/her expected time period, he/she may pay a higher generalized cost and his/her travelling satisfaction will decrease. Therefore, to better describe this dynamic balance status, a satisfactory function of passenger travelling at the  $p$  time period is constructed in this paper, as follows:

$$g_p(Q_p) = \rho \cdot A_p + (1 - \rho) \cdot B_p, \quad (6)$$

where  $\rho$  indicates the passenger's preference of the travelling time period satisfaction and  $0 < \rho < 1$ .

### 3. Bilevel Planning Model

#### 3.1. Lower Layer Planning

(1) *Equilibrium Assignment Conditions of Passenger Flow.* As mentioned above, passengers always expect the maximum satisfaction when travelling. Under the condition that the train operation scheme is set, passengers will choose their travel plans with the maximum satisfaction. It is incompatible with passenger flow equilibrium assignment condition of minimum travelling impedance in UE assignment model. Therefore, according to the feature of value range  $[0, 1]$  in satisfaction function  $g_p(Q_p)$ , the paper constructs an Impedance Function suitable for lower layer planning in this paper, as shown in

$$y_p(Q_p) = 1 - g_p(Q_p). \quad (7)$$

The value range of  $y_p(Q_p)$  is  $[0, 1]$  and minimum of  $y_p(Q_p)$  equals maximum of  $g_p(Q_p)$ . According to utility theory, among all the optional travel plans, the impedances of all the plans chosen by passengers are the same and not greater than those of the unselected plans, as shown in

$$\begin{aligned} y_p &= y_{\min} & \text{if } Q_p > 0 \\ y_p &> y_{\min} & \text{if } Q_p = 0, \end{aligned} \quad (8)$$

where  $y_p$  is the travel impedance of each passenger in the travelling time period  $p$ ;  $y_{\min}$  is the travel impedance under balanced status.

(2) *Equilibrium Assignment Model.* The paper has done an equivalent exchange to the UE assignment model proposed

by Beckmann. The following formula is the objective function:

$$\min Z(k) = \sum_{p \in P} \int_0^{h_p} y_p(k) \cdot dk, \quad (9)$$

where the upper limit  $h_p$  of integral indicates the segment flow on the route  $p$ .

Formula (10) is the sum of passenger flow on all routes. This constraint indicates conservation relation between the route flow and total passenger flow demand:

$$\sum_{p \in P} Q_p = Q. \quad (10)$$

The following formula ensures that all the route flows are positive, thus satisfying the equilibrium assignment condition of passenger flow expressed in formula (9):

$$Q_p > 0. \quad (11)$$

The following formula is the relationship of segment flow and route flow:

$$h_p = Q_p. \quad (12)$$

This paper constructs passenger travel network, and every travelling time period is one route, with total of  $m$  routes. One route only has one segment. Therefore, segment flow  $h_p$  is the route flow  $Q_p$ .

### 3.2. Upper Layer Planning

(1) *Objective Function.* The objective of upper layer planning is to maximize the economic benefits of the railway transportation enterprises, as follows:

$$\max E = \sum_{p \in P} (C \cdot Q_p - W \cdot X_p), \quad (13)$$

where  $X_p$  is the quantity of trains operated in the travelling time period  $p$ ;  $W$  is the operation cost of a single train;  $C \cdot Q_p$  is the ticket revenue;  $W \cdot X_p$  is the operation cost for all trains in the travelling time period  $p$ .

(2) *Dispatching Capacity Constraints of the Station.* On the intercity railway, section carrying capacity is relatively sufficient comparing with the transport demand. However, the dispatching capacities of peak-hour time periods may not meet the transport demand of that time period. Therefore, dispatching capacity constraints should be considered when developing the upper layer planning model, as shown in

$$X_p \leq n_p, \quad (14)$$

where  $n_p$  is the dispatching capacity of the travelling time period  $p$ .

(3) *Constraint of Passenger Flow Demand.* As mentioned above, passenger flow has obvious variability, and the passenger flow demand may be greater than the dispatching capacity

during some time periods of the day. But the transport capacity provided by all the trains operating throughout the day will be greater than the sum of passenger flow demands at each time period of the day, as shown in

$$\sum_{p \in P} D \cdot X_p \geq \sum_{p \in P} M_p, \quad (15)$$

where  $D$  is the maximum seats of a train and  $\sum_{p \in P} D \cdot X_p$  shows the seating capacity of all trains provided by the railway department during one day.

(4) *Passenger Flow Constraints of Each Time Period.* Consider

$$D \cdot X_p \geq Q_p. \quad (16)$$

$D \cdot X_p$  shows that the seating capacity provided by the railway department in the travelling time period  $p$  should be able to satisfy the passenger flow  $Q_p$  of that time period.

(5) *Train Overload Constraint.* During passenger flow peak hours, standing ticket is allowed to sell on the intercity railway. For example, the standing tickets have reached 20% during peak hours on Beijing-Tianjin Intercity Railway. Therefore, this paper proposes train overload constraint, as shown in

$$\frac{Q_p \bmod D}{D \cdot X_p} \leq \pi, \quad (17)$$

where mod is the calculation of dividing modulus and keeping remainder and  $\pi$  is the maximum overload coefficient of the train.

(6) *Nonnegativity and Integer Constraint.* Consider

$$X_p \geq 0 \text{ where } X_p \text{ is an integer.} \quad (18)$$

3.3. *Associations Relationship between Upper Layer Planning and Lower Layer Planning.* Lower layer planning satisfies the UE flow distribution under the Wardrop equilibrium assignment condition and could generate the passenger flow distribution scheme  $Q_p$  according to the passenger flow equilibrium assignment condition. Upper layer planning is the optimized model aiming at maximizing the railway operational benefit and can obtain the train operation scheme  $X_p$  through optimizing calculation.

Upper layer planning and lower layer planning are not isolated but correlated with each other. Based on the train operation scheme  $X_p$  defined by upper layer planning, the correlation parameters of Impedance Function for lower layer planning can be obtained as

$$I_p = \frac{1}{X_p}. \quad (19)$$

This paper assumes that trains are evenly dispatched at one time period and each travelling time period equals 1 hour without exception. Therefore,  $1/X_p$  shows the train dispatching interval of the time period  $p$ .

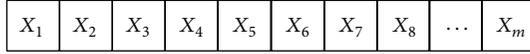


FIGURE 1: Chromosome structure.

According to the passenger flow distribution scheme  $Q_p$  of lower layer planning, the ticket revenue  $\sum_{p \in P} C \cdot Q_p$  of upper layer planning and passenger flow constraint conditions of each time period can be obtained.

Through the nested calculation of the upper layer and lower layer planning, this paper proposes the equilibrium relationship between upper layer optimized scheme  $X_p$  and passenger flow distribution scheme  $Q_p$  of lower layer, as follows:

$$X_p = \text{ceil}\left(\frac{Q_p}{D}\right) + \theta, \quad (20)$$

$$\theta = \begin{cases} 0 & \text{if } \frac{Q_p \bmod D}{D} < \eta, \\ 1 & \text{if } \frac{Q_p \bmod D}{D} \geq \eta, \end{cases} \quad (21)$$

where  $\text{ceil}(j)$  is roundup function and  $\eta$  is the minimum occupancy rate of the train;  $\theta$  is variable between 0 and 1, the value of which is calculated according to formula (21).

## 4. Algorithm Design

The model established in this paper is a bilevel planning problem of nonlinear mixed integer and is widely acknowledged as one of the optimization problems that are extremely difficult to solve. This paper adopts a genetic algorithm to get the solution of the model. This algorithm has better global convergence and is able to search for satisfactory solution in a short time, the core concept of which is to keep the high fitness of the individual to the next generation while eliminating the low fitness ones, based on the principle of natural evolution ‘‘Survival of the fittest.’’

### 4.1. Genetic Algorithm Design

(1) *Chromosome Coding.* Chromosome is made up of  $m$  gene segments which indicate  $m$  travelling time periods, respectively. Integer coding is used for chromosome coding. The position of each chromosome indicates the quantity of dispatched trains in the corresponding time period. The structure is as shown in Figure 1.

(2) *Fitness Function.* In the genetic algorithm, whether the individual can be kept to the next generation is determined by its fitness. The higher the individual fitness is, the bigger possibility it gets to be transmitted to the next generation; conversely the possibility is smaller. This paper adopts the following formula as a fitness function:

$$\text{fitness} = E. \quad (22)$$

In the formula,  $E$  indicates the target value of chromosome.

(3) *Genetic Operation.* Genetic operation (selection operation, crossover operation, and mutation operation) is the core step of genetic algorithm, and it can produce a new generation through the genetic operation to the population. This paper adopts the improved crossover mutation method in literature [11].

### 4.2. Algorithm Process

*Step 1 (initialization).* According to the above described chromosome coding regulations and formulae (14)–(18), randomly generate the initial population of feasible solution with the scale of popsize; set optimal goal  $E^* = 0$  of the upper layer planning; initialize operation scheme  $\mathbf{WP}^*$  to zero vector; proposal inspection set  $\Psi$  is an empty set; the number of iterations is  $t = 1$ .

*Step 2 (genetic operation).* Based on the genetic algorithm designed in Section 4.1, conduct the selection, crossover, and mutation operation to the initial population, and search for feasible solution (operation scheme)  $\mathbf{WP}^t$  with the highest fitness in the current population; the number of iterations is  $t = t + 1$ .

*Step 3 (checking).* If  $\mathbf{WP}^t \notin \Psi$ , then add it into check set  $\Psi$  and go to Step 4; otherwise, go to Step 2.

*Step 4 (user equilibrium assignment).* For the current operation scheme  $\mathbf{WP}^t$ , use the Frank-Wolfe algorithm to get solution for lower layer planning, and obtain  $Q_{pt}$  in conformity with the Wardrop user equilibrium criterion.

*Step 5.* Bring the current operation scheme  $\mathbf{WP}^t$  and passenger flow demand  $Q_{pt}$  into the upper layer objective function, and calculate the objective function  $E$ ; if  $E > E^*$ , then let  $E^* = E$  and  $\mathbf{WP}^* = \mathbf{WP}^t$ .

*Step 6 (termination of inspection).* If the number of iterations  $t$  is greater than the upper limits  $G$ , then generate the optimal solution  $\mathbf{WP}^*$ ; otherwise, go to Step 2.

## 5. Numerical Example

The paper takes the optimization of train operation scheme on Beijing-Tianjin Intercity Railway as an example to verify the model and algorithm presented in this paper.

### 5.1. Parameter Value

(1) *Division of Time Period.* The paper divides the service time (7:00–23:00) of Beijing-Tianjin Intercity Railway into 17 intervals, on the basis of one hour per interval.

(2) *Passenger Travel Demands.* The paper uses the official forecast data of the Beijing Railway Bureau. The average daily passengers would reach 117,400 by 2015 on Beijing-Tianjin Intercity Railway; that is,  $M = 117400$ . According to passenger flow survey, the passenger travel demand proportion of different time periods on Beijing-Tianjin Intercity Railway is

TABLE 1: Passenger travel demands of different time periods on Beijing-Tianjin Intercity Railway.

$p$	1	2	3	4	5	6	7	8	9
$M_p$	2980	8939	11919	11919	5959	2384	1788	1192	5363
$p$	10	11	12	13	14	15	16	17	
$M_p$	4768	5959	9535	10727	11919	10131	7747	4172	

TABLE 2: Calculation result.

$p$	$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
	$Q_p$	$X_p$	$\tau_p$	$Q_p$	$X_p$	$\tau_p$	$Q_p$	$X_p$	$\tau_p$
1	6212	7	0.96	6763	7	1.04	6084	7	0.94
2	6212	7	0.96	8846	10	0.95	8196	9	0.98
3	5218	6	0.94	6763	7	1.04	10014	11	0.98
4	8324	9	1	7310	8	0.98	8068	9	0.97
5	7454	8	1	7586	8	1.02	7301	8	0.98
6	7827	8	1.05	6043	7	0.93	5635	6	1.01
7	6833	7	1.05	6783	7	1.04	3458	4	0.93
8	5466	6	0.98	6783	7	1.04	7556	8	1.02
9	7206	8	0.97	5566	6	1	7684	8	1.04
10	7206	8	0.97	8390	9	1	7300	8	0.98
11	8199	9	0.98	8866	10	0.96	6403	7	0.99
12	5218	6	0.94	3923	4	1.06	6019	6	1.08
13	9939	11	0.97	7023	8	0.95	8196	9	0.98
14	6460	7	0.99	6773	7	1.04	8837	10	0.95
15	4100	4	1.1	6773	7	1.04	6916	7	1.06
16	9314	10	1	6470	7	1	4354	5	0.94
17	6212	7	0.96	6739	7	1.04	5379	6	0.97
Total	117400	128	/	117400	126	/	117400	128	/

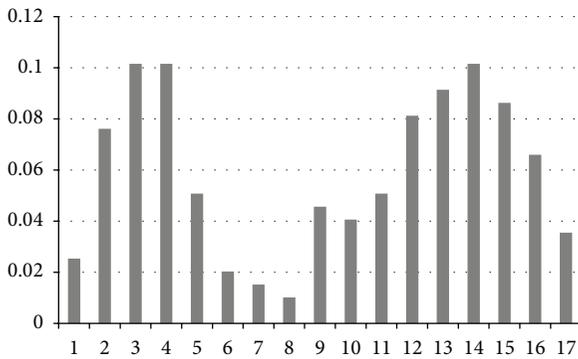


FIGURE 2: Passenger flow distribution of different time periods on Beijing-Tianjin Intercity Railway.

as shown in Figure 2. Therefore, the passenger travel demand  $M_p$  in each time period is as shown in Table 1.

(3) *Relevant Operation Parameters of Beijing-Tianjin Intercity Railway.* Relevant operation parameters are set as [8, 9]:  $C$  is 57.25 yuan,  $D$  is 928 person,  $W$  is 18000 yuan/train,  $N$  is 5000 person,  $\lambda$  is 21.3 yuan,  $\pi$  is 0.2,  $\eta$  is 0.6,  $n_p = \{10, 12, 12, 10, 10, 10, 10, 10, 10, 10, 12, 12, 12, 10, 10, 10, 10\}$ , and  $p \in P$ .

(4) *Algorithm Parameter Settings.* According to the algorithm designed in the paper, the population size  $\text{popsize} = 100$ ; obtaining crossover probability  $P_c = 0.6$  and mutation probability  $P_m = 0.4$  based on experience, the maximum stagnation iteration number is 15 and the maximum iteration number is 500.

5.2. *Calculation Result and Analysis.* This paper considers three kinds of situations,  $\rho = 0.25$ ,  $\rho = 0.5$ , and  $\rho = 0.75$ , and calculates the passenger flow  $Q_p$ , quantity of trains operated  $X_p$ , and average occupancy rate  $\tau_p$  in each time period, respectively. The calculation result is as shown in Table 2.

The paper obtains the following information based on the calculation data in Table 2. (1) Quantity of trains operated  $X_p$  in each time period is less than dispatching capacity  $n_p$ , which meets the train operation condition. (2) Average train occupancy rate  $\tau_p$  of each time period is greater than the minimum occupancy rate  $\eta$  and  $\tau_p - 1$  is less than the maximum overload rate  $\pi$ . (3) When  $\rho$  increases, passenger's requirements of travelling time period satisfaction rise and passenger flow  $Q_p$  shows obvious peaks and troughs in each time period. When  $\rho = 0.75$ , passenger flow  $Q_p$  changing trend of each time period is consistent with transport demand  $M_p$  to the greatest extent, as shown in Figure 3. (4) When  $\rho = 0.5$ , the total number of dispatched trains in all time

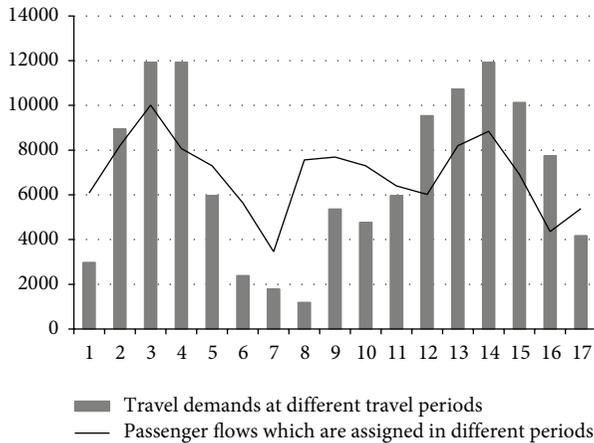


FIGURE 3: At the point of  $\rho = 0.75$ , comparison of changing trend for  $Q_p$  and  $M_p$  in each time period.

periods is 126; when  $\rho = 0.75$  and  $\rho = 0.25$ , the total number of dispatched trains in all time periods is 128.

## 6. Conclusions

The paper applies the UE equilibrium assignment theory to the decision of travel plans for intercity railway passengers and proposes the bilevel planning model of train operation scheme on intercity rail lines. Aiming at maximizing passenger's average travelling satisfaction, the paper constructs the passenger's travel impedance function and uses this function as passenger flow equilibrium assignment condition for lower layer planning to conduct assignment; meanwhile, it takes the maximum benefit of railway transport enterprises as target to make optimization. Through optimizing calculation, it obtains the passenger flow of each time period and the train operation scheme. By analyzing the calculation results, the model and algorithm proposed in this paper can effectively optimize the train operation scheme of intercity rail lines. Next, the paper will conduct a deep research on passenger travel choice under the condition of time-based pricing.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

The work described in the paper was supported by the National Natural Science Foundation of China under Grant no. 71261014.

## References

[1] B. Mandel, M. Gaudry, and W. Rothengatter, "A disaggregate box-cox logit mode choice model of intercity passenger travel in Germany and its implications for high-speed rail demand

forecasts," *Annals of Regional Science*, vol. 31, no. 2, pp. 99–120, 1997.

- [2] C. J. Zhao, *Study on Theory and Application of Inter-City Rail Network Planning*, China Academy of Railway Science, Beijing, China, 2009.
- [3] E. Cascetta and P. Coppola, "An elastic demand schedule-based multimodal assignment model for the simulation of high speed rail (HSR) systems," *EURO Journal on Transportation and Logistics*, vol. 1, no. 1-2, pp. 3–27, 2012.
- [4] F. Shi, Z. C. Huang, W. L. Zhou, and L. B. Deng, "Optimization on departure time distribution of passenger trains based on user equilibrium analysis," *Journal of Railways Science and Engineering*, vol. 5, no. 6, pp. 69–75, 2008.
- [5] H. Q. Peng and Y. J. Zhu, "Intercity train operation schemes based on passenger flow dynamic assignment," *Journal of Transportation Systems Engineering and Information Technology*, vol. 13, no. 1, pp. 111–117, 2013.
- [6] H. M. Niu and X. S. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transportation Research C*, vol. 36, pp. 212–230, 2013.
- [7] Y.-Q. He, H.-Z. Zhang, B.-H. Mao, and T.-S. Chen, "Multiobjective bi-level programming model of making train working plan for passenger-only line," *Journal of the China Railway Society*, vol. 28, no. 5, pp. 6–10, 2006.
- [8] F. Shi, L. Deng, and L. Huo, "Bi-level programming model and algorithm of passenger train operation plan," *China Railway Science*, vol. 28, no. 3, pp. 110–116, 2007.
- [9] Y.-H. Chang, C.-H. Yeh, and C.-C. Shen, "A multiobjective model for passenger train services planning: application to Taiwan's high-speed rail line," *Transportation Research B*, vol. 34, no. 2, pp. 91–106, 2000.
- [10] Z. Huang and H. Niu, "Study on the train operation optimization of passenger dedicated lines based on satisfaction," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 451201, 11 pages, 2012.
- [11] G. Luo, E. L. Liu, and J. Wang, "Resource planning optimization in network schedule using genetic algorithms," *Journal of Tianjin University*, vol. 32, no. 2, pp. 179–183, 2004.

## Research Article

# Capacity Estimation for On-Ramp Merging Section of Urban Expressway Based on Time Headway Loss

Xing-jian Xue,<sup>1,2,3</sup> Feng Shi,<sup>1</sup> and Qun Chen<sup>1</sup>

<sup>1</sup> School of Transportation Central South University, Changsha 410075, China

<sup>2</sup> Hunan Province University Key Laboratory of Bridge Engineering (Changsha University of Science & Technology), Changsha 410000, China

<sup>3</sup> School of Transportation and Logistics Central South University of Forestry and Technology, Changsha 410004, China

Correspondence should be addressed to Qun Chen; chenqun631@csu.edu.cn

Received 25 September 2013; Revised 31 December 2013; Accepted 1 January 2014; Published 20 February 2014

Academic Editor: Huimin Niu

Copyright © 2014 Xing-jian Xue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a model for estimating capacity of on-ramp merging section of urban expressway based on dynamics and gap acceptance theory, considering lane-changing processes and time headway loss. Survey data were collected from on-ramp merging sections of shanghai urban expressway system and showed that capacity drop of on-ramp merging section is caused by drivers' lane-changing which may lead to unsteady speed of vehicles and so prolonged time headway compared to the minimum time headway corresponding to the maximum capacity. Three parameters (optimal time headway, time headway loss, and interference quantity of lane-changing) are given and a methodology by accumulating time headway loss due to lane-changing is developed to estimate the capacity drop. Results' comparisons between real data and microsimulation of on-ramp merging sections and sensitivity analysis show that the proposed model can produce reliable and accurate results. This study also reveals that ramp flow and the difference between the optimal speed and the lane-changing speed of fleet have a great impact on capacity drop. This study is beneficial to evaluate congestion levels, to understand complex traffic phenomena, and so to find efficient solutions.

## 1. Introduction

Traffic congestion causes a huge disadvantage, such as waste of time and fuel as well as pollution. On urban expressways, congestion is usually induced by various kinds of bottleneck, among which the on-ramp merging section is an important position and has been widely studied. To enhance the capacity of an expressway with an on-ramp merging section, various kinds of methods have been proposed, in which an accurate traffic model that can efficiently estimate the capacity of on-ramp is most important while evaluating congestion levels, understanding complicated traffic phenomena, putting forward reasonable design scheme of infrastructure, and developing efficient operation strategies.

In the past decades, continuous effort has been devoted to the study of capacity of on-ramp merging section. Many researchers have tried to identify the causes of capacity drop at on-ramp merging section that seems to be linked to

flow interference, which are created by vehicle lane-changing maneuvers [1–6].

Several models have been proposed to account for capacity. The most commonly studied model is probabilistic model based on gap acceptance theory. The gap acceptance theory was put forward by Drew [7]. Westphal concluded that the on-ramp capacity decided by outmost lane flow [8]. Lertworawanich and Elefteriadou developed a new gap acceptance model through definition expressway capacity as function of ruin probability [9]. Kim and Son promote a new model of on-ramp capacity through making time headway obey different flow range [10], but most of these models are hard to implement and to calibrate in practice [2, 11]. However, the development of new tools now allows having access to more refined information and more rapid calculations [10]. In contrast, the method put forward by HCM based on regression analysis not only is simple but also has very strong practicability [10]; its disadvantage is high

cost of data investigation and small scope of application. In addition, there are traffic wave capacity model, cell automaton capacity model, and so forth [10]. Recent studies have shown that the capacity of expressway is not immutable and frozen; many researchers suggest that capacity can only truly be defined as a function of breakdown probability [12, 13].

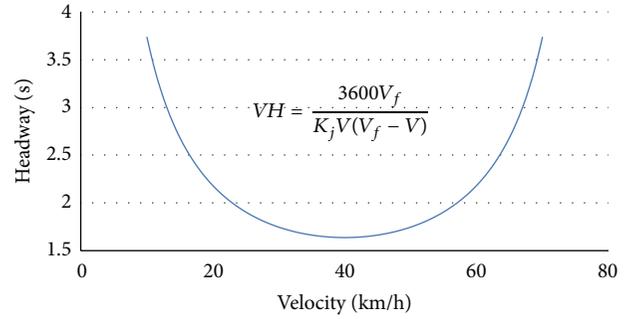
The overall objective of this study is to develop a new on-ramp merging section capacity estimation model of urban expressway, which considers lane-changing processes, vehicle speed of on-ramp, vehicle flow, and time headway distribution. At first, a new modeling method about capacity is discussed and some new parameters were proposed. Then, this paper develops a series of hypotheses about capacity model, and the study site and traffic data in detail have been described. Section 3 builds the capacity model about on-ramp merging section of urban expressway, and the parameters were calibrated. In the final section, the data obtained from the capacity model are compared with field-observed capacities and microsimulation result, and sensitivity analyses show that our model is robust; the implications of the research results are discussed in different fields.

## 2. Idea of Mathematical Modeling and Its Theoretical Explanation

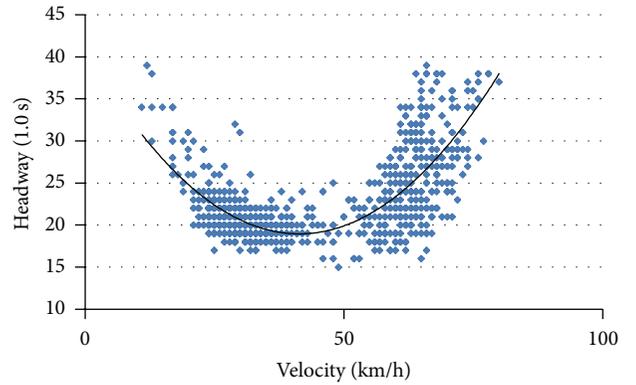
The traffic flow is inversely proportional to the time headway of vehicles. Without lane-changing, time headway is related to some factors such as speed, traffic flow interference, and the reaction time of drivers. The time headway reaches a minimum value when capacity reaches a maximum value. The minimum time headway is just the optimal headway, symbolized as  $VH_o$ , and the corresponding speed is the optimal speed, symbolized as  $V_o$ .

The relationship between the flow and the velocity can be derived from the Greenshields model, and its curve of relationship is shown in Figure 1(a). The figure shows that the optimal speed corresponds to the minimum time headway. However, because in most cases the real situation cannot correspond to the results obtained from the model well, the relation between speed and time headway should be derived from measured data. In this paper, the scatter diagram (shown in Figure 1(b)) about the relation of speed and headway was obtained from using one week 5 min loop-detector data at Shanghai expressway system. The position of data investigation is the inner lane of Yan'an road of the Shanghai expressway system because here the traffic flow has the most stable state and the least merging interference. The survey data show that the capacity can reach a maximum value (2050 pcu/h) that corresponds to the minimum time headway (1.8 s) when the speed keeps about 45 km/h, as shown by Figure 1(b); the result is consistent with the actual situation. Therefore, this paper defines 2000 pcu/h as the capacity per one lane, the 1.8 s as the optimal headway, and 45 km/h as the optimal speed.

On-ramp merging sections usually become the bottleneck of urban expressway because headway cannot keep an optimal value; namely, the time headway is prolonged due to unsteady traffic flow caused by lane-changing of vehicles



(a) The model of Greenshields



(b) The fitting of the measured data

FIGURE 1: The relation between time headway and vehicle speed.

and so frequently variable speed. The prolonged part of time headway from lane-changing can be used for more vehicles passing and is called as the single loss of time headway, symbolized as  $T_l$ . The operational state of vehicle at on-ramp merging section is shown in Figure 2(a). When the vehicle running on the most medial lane passes the on-ramp merging section with a speed of  $V_o$ , whose corresponding time headway is  $VH_o$ , the time headway of vehicles running on the outermost lane would prolong due to interferes from flows merging. The fleet of different quantity scale would be merging to mainline according to gap size between vehicles running on the most lateral lane. For example, fleet 1 consists of one car; fleet 2 consists of two cars, as shown in the Figure 2(b). The speed of vehicles running on the mainline behind fleet 1 is forced to slow down to  $V_1$  because of the interference of fleet 1's lane-changing to the mainline with this speed. So, a larger headway  $VH_1$  appears between vehicle 11 of fleet 1 as head car and vehicle 2. It is the loss of travel time ( $t_l$ ) for lane-changing of fleet 1 which is used for speed 1 arising to speed 2 without considering other factors such as drivers' reaction time and internal interference, the same reason for analysis of fleet 2. The number of merging fleet in one unit of time is called as the interference quantity of lane-changing, denoted by  $q_{rg}$ . The interference quantity of lane-changing is not equal to the vehicle number of lane-changing; it is equal to the number of major road gap that is greater than critical gap of lane-changing. For example, there are 10 ramp vehicles; if each gap of major road can only merge one

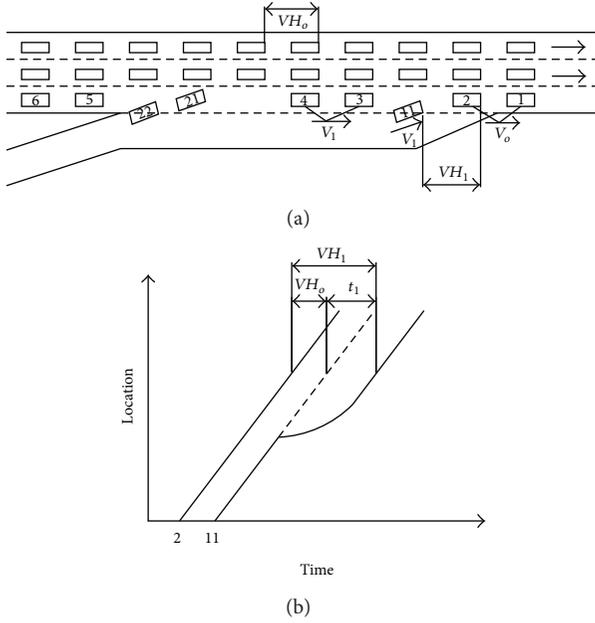


FIGURE 2: Traffic flows at on-ramp merging section.

vehicle, then the interference quantity of lane-changing is 10; if each gap of major road can merge two vehicles, then the interference quantity of lane-changing is 5. Because there will only be one interference in a gap of major road traffic flow, do not change with the scale of the merge fleet. The product of  $q_{rg}$  and  $t_l$  is called as the headway loss per unit time. The difference between unit time and the headway loss is called as the available travel time. The capacity of lateral lane can be obtained if the flows of ramp and upstream mainline of lateral lane have been known. Finally, the capacity of on-ramp merging section can be obtained from accumulation of all of the lanes of the expressway.

### 3. The Capacity Model for On-Ramp Merging Section Based on Time Headway Loss

**3.1. The Description of Parameters and Variables.**  $a$  is the acceleration of lane-changing fleet;  $C$  is the capacity of on-ramp merging section;  $C_I$  is the capacity of inside lane;  $C_o$  is the capacity of lateral lane of mainline;  $i$  is the vehicle number of lane-changing fleet,  $i \geq 1$ ;  $l$  is the order of Erlang distribution;  $m$  is the lane number of mainline;  $n$  is the maximum number of vehicles in lane-changing fleet;  $p(t_{vh} > t_c)$  is the probability of  $t_{vh}$  being greater than  $t_c$ ;  $p[t_c + it_f > t > t_c + (i-1)t_f]$  is the probability of headway  $t$  between  $t_c + it_f$  and  $t_c + (i-1)t_f$ ;  $q_m$  is the rate of flow in lateral lane of mainline upstream;  $q_r$  is the vehicle number of lane-changing;  $q_{rg}$  is the interference quantity of lane-changing;  $T_a$  is the time for vehicles' acceleration from lane-changing speed to the optimal speed;  $T_b$  is the basic loss headway, namely, the loss headway due to driver reaction time and internal interference of fleet;  $T_l$  is the loss headway per lane-changing of the fleet;  $t_{vh}$  is time-headway;  $t_c$  is critical gap of lane-changing;  $t_f$  is the headway of car-following;  $t_0$  is the minimum headway;

$V_o$  is optimal speed;  $V_r$  is the speed of vehicles running on the on-ramp;  $V_c$  is the difference between the optimal speed and the lane-changing speed of fleet (speed difference of lane-changing for short);  $VH_o$  is optimal headway;  $\lambda$  is arriving rate of vehicles.

**3.2. Basic Hypothesis.** The model is based on some assumptions, including the following.

- (1) The same type of vehicles has the same critical gap.
- (2) The speed is consistent when these vehicles change lanes.
- (3) The vehicle running on the ramp could change to mainline only if the gap of mainline is greater than the critical gap ( $t_c$ ).
- (4) The effects of each lane-changing caused by interference will not spread to the upstream.
- (5) The on-ramp traffic demand is saturated.

**3.3. Model Construction.** The capacity model of on-ramp merging section is shown in formula (1). It consists of two parts. The first part is the capacity of lateral lane of mainline, and its interference by lane-changing vehicle is shown in formula (2). The second part is the capacity of inside lane which is not disturbed by lane-changing vehicles, as shown in formula (3). Consider

$$C = C_o + C_I, \quad (1)$$

$$C_o = \frac{3600 - T_l q_{rg}}{VH_o}, \quad (2)$$

$$C_I = \frac{3600}{VH_o} (m - 1). \quad (3)$$

Parameter  $T_l$  consists of the acceleration time of driving on-ramp ( $T_a$ ) and the loss headway due to internal interference of fleet ( $T_b$ ), as shown in formula (4). The size of acceleration time ( $T_a$ ) is related to the speed difference between  $V_o$  and  $V_r$ , as well as the acceleration of fleet ( $a$ ) as shown in Figure 2(b). Their relationship is shown in formula (5). Consider

$$T_l = T_a + T_b, \quad (4)$$

$$T_a = \frac{(V_o - V_r)}{a}. \quad (5)$$

Parameter  $q_{rg}$  refers to the interference number due to lane-changing, and it is related to flow of lateral lane of expressway. Its meaning is different from  $q_r$ . The gap size distribution in different flow conditions is also different. When the flow is small, the size of gap is so small that the merging fleet is composed of only one car in most cases; thus the gap between  $q_{rg}$  and  $q_r$  is small. On the contrary, the gap is larger and the merging fleet consists of more vehicles. The  $q_{rg}$  is the cumulative quantity of lane-changing fleet of different scale, and its model is shown in formula (6). Its depends on

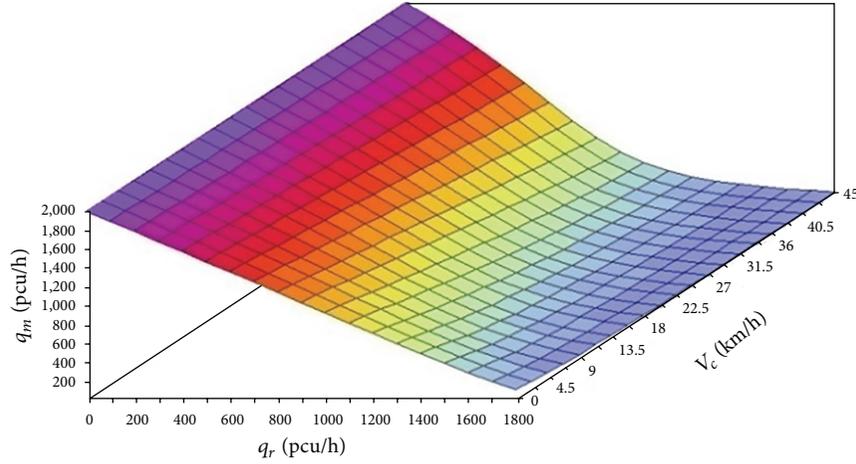


FIGURE 3: The relations between  $q_r$ ,  $V_c$ , and  $q_m$ .

$t_c$  and  $t_f$  about how large-scale fleet may merge to mainline in some headway. Consider

$$q_{rg} = \sum_{i=1}^n \frac{p[t_c + it_f > t_{vh} > t_c + (i-1)t_f] q_r}{p(t_{vh} > t_c) i}. \quad (6)$$

Bringing formulas (5)~(8) into formula (4), then the formula of on-ramp merging section capacity of urban expressway based on time headway loss is shown in the following formula:

$$C = \left( \left( 3600 - \left( \frac{|V_o - V_r|}{a} + T_b \right) \times \sum_{i=1}^n \frac{p[t_c + it_f > t_{vh} > t_c + (i-1)t_f] q_r}{p(t_{vh} > t_c) i} \right) \times (VH_o)^{-1} \right) + \frac{3600}{VH_o} (m-1). \quad (7)$$

The headway of lateral lane obeys different kinds of distributions with the flows changing. The Erlang distribution is suitable for describing headway distribution according to different flow conditions because it can describe different degree of flow conditions from unblocked to congestion by adjusting the order  $L$ , as shown in formula (8). The Erlang distribution about order 1 is the negative exponential distribution, and the headway will keep equal when the value of order  $L$  is very large. The research of the literature showed that the flow range of main lane in less than 700 pcu/h, 700~1500 pcu/h, and more than 1500 pcu/h was applied to the Erlang distribution of order 1, 2, 3 [14]. The usual value range of  $t_0$  was 11.5 s. Consider

$$p(t_{vh} \geq t) = \sum_{i=0}^{l-1} (\lambda t)^i \frac{e^{-\lambda(t-t_0)}}{i!}. \quad (8)$$

### 3.4. Parameter Calibration

- (1)  $t_c$  and  $t_f$ : the literature research showed that the average value of  $t_c$  is 3.0 s and  $t_f$  is 1.5 s according to the survey based on 200 observation samples in some on-ramp merging sections of Beijing [15].
- (2)  $a$ : the  $2 \text{ m/s}^2$  is a suitable value according to the mechanical properties of cars.
- (3)  $T_b$ : the 0.5 s is a suitable value according to survey data and microsimulation analysis.
- (4)  $n$ : although the vehicles in fleet per lane-changing may be more when the flow in the mainline's lateral lane is very small, the suggested value of  $n$  is 4~8 for computational simplicity and accuracy.

## 4. Model Calculation

The relations between  $q_r$ ,  $V_c$ ,  $q_m$ , and  $C_o$  have been obtained by the model programming and solving using LM (Levenberg-Marquardt) algorithm, as shown in Figures 3 and 4 and Tables 1 and 2. As shown in Figure 3, the relation between  $q_r$  and  $q_m$  is linear, and the relation will be converted into nonlinear with the value of  $V_c$  increasing. As shown in Figure 4, the main body of flow has changed from the vehicles driving on mainline to those driving on on-ramp with the increase of on-ramp vehicles. The capacity also showed a rising trend after the first drop. It showed that ramp traffic flow had a negative impact on the capacity of on-ramp merging section of expressway, and this negative effect was proportional to  $V_c$ .

As shown in Table 1, the capacity reached the lowest when the value of  $q_m$  takes a value between 300~430 pcu/h and  $q_r$  takes a value between 1000~1500 pcu/h with the increase of  $V_c$ . Table 2 shows that the maximum loss of capacity can reach 1/3 of the total. This shows that it is very important that vehicles of ramp have enough acceleration space.

This model shows that the critical parameter for the capacity of on-ramp merging section is the  $q_r$  and  $V_c$ . It is an effective way reducing the value of these two factors to

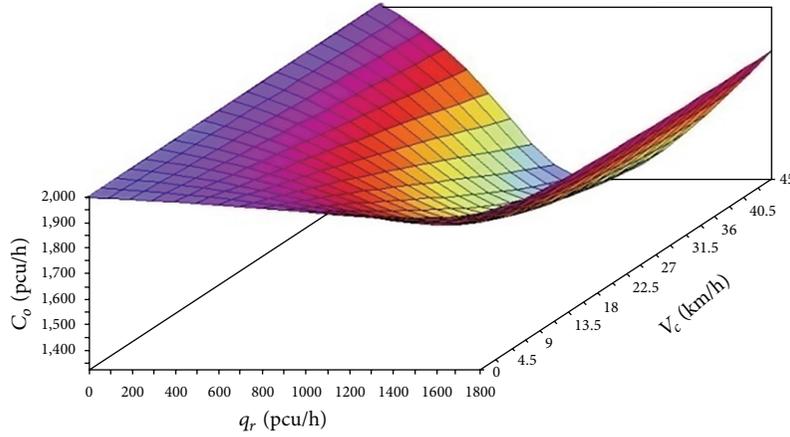


FIGURE 4: The relations between  $q_r$ ,  $V_c$ , and  $C_o$ .

TABLE 1: The relations between  $q_r$ ,  $V_c$ , and  $q_m$ .

$q_m$ (pcu/h)	$V_c$ (km/h)										
	0	4.5	9	13.5	18	22.5	27	31.5	36	40.5	45
200	1793	1783	1774	1764	1754	1745	1735	1725	1715	1704	1694
400	1583	1562	1540	1518	1495	1472	1448	1423	1398	1373	1346
600	1372	1336	1299	1261	1221	1180	1138	1094	1050	1004	957
800	1159	1106	1051	995	936	876	815	755	695	637	583
1000	945	874	802	728	655	586	520	461	409	363	325
1200	731	645	561	482	412	352	303	263	230	204	183
1400	521	429	349	284	234	197	168	146	129	115	104
1600	320	241	185	147	120	101	87	76	68	61	55
1800	141	98	73	58	48	41	35	31	28	25	23
$T_l$ (s)	0.00	0.63	1.25	1.88	2.50	3.13	3.75	4.38	5.00	5.63	6.25

TABLE 2: The relations between  $q_r$ ,  $V_c$ , and  $C_o$ .

$C_o$ (pcu/h)	$V_c$ (km/h)										
	0	4.5	9	13.5	18	22.5	27	31.5	36	40.5	45
200	1993	1983	1974	1964	1954	1945	1935	1925	1915	1904	1894
400	1983	1962	1940	1918	1895	1872	1848	1823	1798	1773	1746
600	1972	1936	1899	1861	1821	1780	1738	1694	1650	1604	1557
800	1959	1906	1851	1795	1736	1676	1615	1555	1495	1437	1383
1000	1945	1874	1802	1728	1655	1586	1520	1461	1409	1363	1325
1200	1931	1845	1761	1682	1612	1552	1503	1463	1430	1404	1383
1400	1921	1829	1749	1684	1634	1597	1568	1546	1529	1515	1504
1600	1920	1841	1785	1747	1720	1701	1687	1676	1668	1661	1655
1800	1941	1898	1873	1858	1848	1841	1835	1831	1828	1825	1823
$T_l$ (s)	0.00	0.63	1.25	1.88	2.50	3.13	3.75	4.38	5.00	5.63	6.25

improve the capacity of on-ramp merging section. There are two points of important significance. First, the value of  $V_c$  can be reduced through setting the acceleration lane longer, but the measures must be taken to control on-ramp flow according to the situation of upstream flow of mainline's lateral lane and avoid the situation that on-ramp vehicles

cannot change from ramp to mainline, which may lead to vehicles queuing on acceleration lane and so the space of acceleration cannot reach the expected. Second, the balance design of lane number of upstream and downstream of merging sections can decrease the value of  $q_r$ , because parts of on-ramp vehicles have no need of lane-changing.

TABLE 3: The data comparison between model and actual measurement.

Serial number	Location	Category	Measured data (pcu/h)	Model calculation (pcu/h)	
				Value	Error
1	Wuyi road	Capacity	3631	3755	3.4%
		The flow of on-ramp	742	742	—
2	Wuyi road	Capacity	3530	3732	5.7%
		The flow of on-ramp	543	543	—
3	Wuning road	Capacity	3761	3613	-3.9%
		The flow of on-ramp	1243	1243	—
4	Wuning road	Capacity	3701	3613	-2.4%
		The flow of on-ramp	1259	1259	—
5	Jinshajiang road	Capacity	5212	5443	-4.1%
		The flow of on-ramp	852	852	—
6	Jinshajiang road	Capacity	5287	5451	-3.0%
		The flow of on-ramp	852	852	—

## 5. Comparison of Computational Results and Sensitivity Analysis

**5.1. Comparison of Computational Results.** In order to test the accuracy of the results of model, this model calculation results are compared with the measured data of Shanghai expressway. These measured data are collected from inner ring of Shanghai urban expressway on-ramp merging section from 8:00 to 10:00, the cross section type of expressway is bidirectional and four lanes, and the length of all the acceleration lane is about 100 meters. The results of comparison are shown in Table 3. The error between model calculation and measured data is within 5% according to the comparison in Table 3. In order to embody the relation between on-ramp flow and the flow of mainline lateral lane, this paper compares these data including VISSIM microsimulation and the model calculation results based on four kinds value of  $V_c$ , because there is not any road but only one lane in reality. The result of comparison show that there is a higher degree of fitting, as shown in Figure 5. This proved that the model has a high accuracy.

**5.2. Sensitivity Analysis.** In order to test whether the applicability of the model is better, sensitivity analysis carried out on the main parameters of the model are necessary. There are two parameters which have great influence on the model:  $VH_o$  and  $V_c$ .

The meaning of sensitivity analysis of  $VH_o$  is analysing the maximum capacity of lateral lane of expressway mainline. The change relation between  $q_r$  and  $C_o$  is shown in Figure 6 when the capacity value of lateral lane is 1800 pcu/h, 2200 pcu/h, and 2400 pcu/h. According to Figure 6, the relation between  $q_r$  and  $C_o$  basic keeps consistent with the increase of capacity. The maximum flow of on-ramp is increasing leading to the maximum drop of capacity with the increase of interference amplitude of on-ramp flow to the capacity of mainline. This is consistent with the actual situation because the aggrandizement of difference between on-ramp traffic flow and mainline traffic flow of expressway. The influence is increasing when vehicle lane is changing.

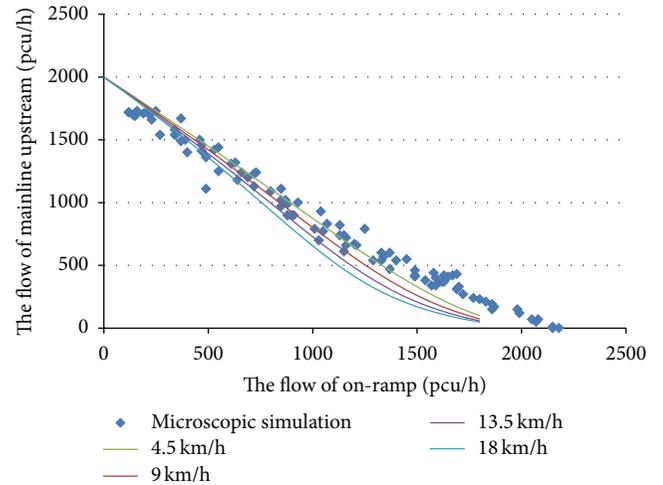


FIGURE 5: The data comparison between model and microscopic simulation of single lane.

The meaning of sensitivity analysis of  $V_c$  expands the range of  $V_c$ , and its meaning to improving the value of the  $V_o$ . The value range of  $V_o$  in this paper is from 45 km/h to 80 km/h, and the corresponding value range of  $V_c$  is from 45 km/h to 80 km/h (the value range from 0 km/h to 45 km/h has analysis in the Figures 3 and 4). The relation between  $q_r$ ,  $q_m$ , and  $C_o$  is shown in Figure 7 after adjustment of  $V_o$ . As shown in Figure 7, the interference of on-ramp vehicle is increasing with the increasing of  $V_c$  and the flow of on-ramp that has great influence is corresponding decreasing with the increasing of  $V_c$ . This is consistent with the actual situation, because the interference of lane-changing behavior of on-ramp vehicle to the flow of mainline is increasing with the increasing of  $V_c$ .

## 6. Conclusions

Upon the completion of the thesis, the following conclusions were drawn.

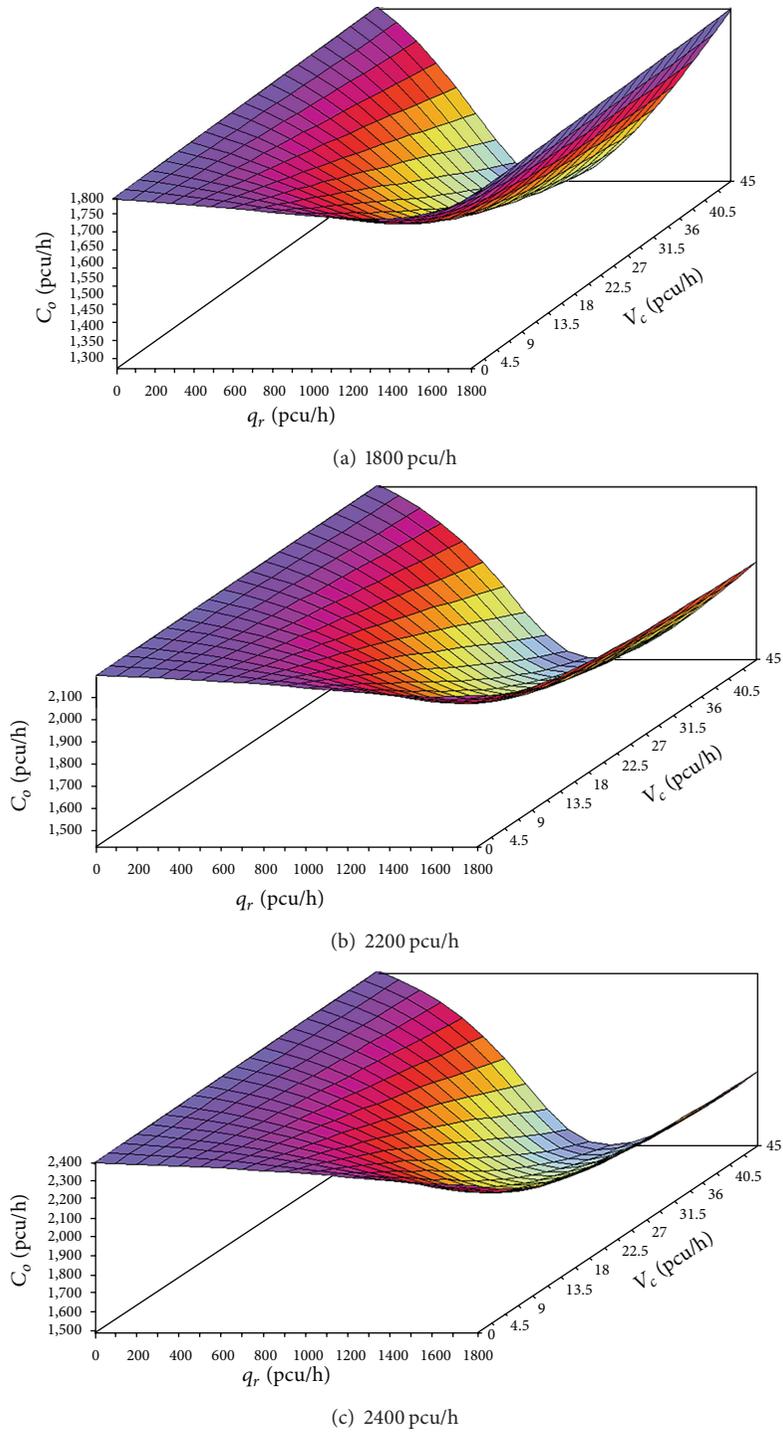
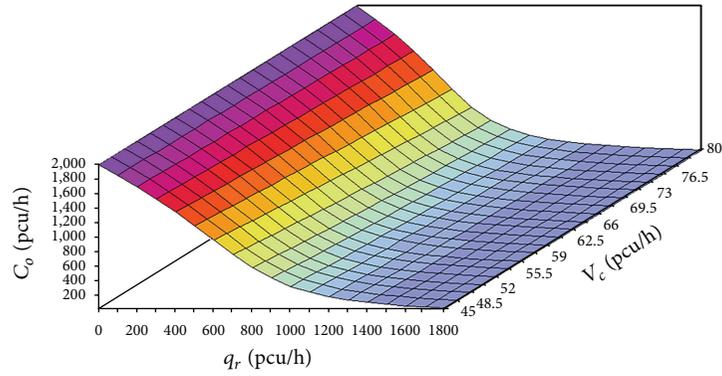
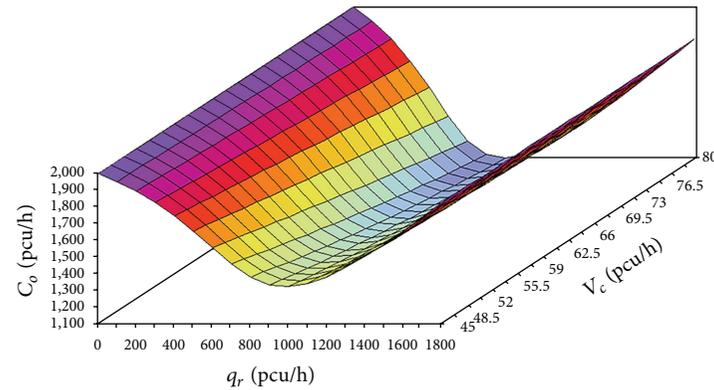


FIGURE 6: The relation diagram between  $q_r$ ,  $V_c$ , and  $C_o$  about different  $C_o$  values.

- (1) The capacity of on-ramp merging section is still a theoretical notion that lacks an unambiguous capacity value; this paper is to address this issue by studying its characteristics under different flow conditions using one week 5 min loop-detector data at Shanghai expressway system.
- (2) This study puts forward three parameters: optimal time headway, time headway loss, and interference quantity of lane-changing. The capacity models for on-ramp merging section based on time headway loss were developed as a function of volume ratios from both of the mainline and the ramp, the gap between

(a) The relation diagram between  $q_r$  and  $q_m$ (b) The relation diagram between  $q_r$  and  $C_o$ FIGURE 7: The relation between  $q_r$ ,  $q_m$ , and  $C_o$  after  $V_c$  adjustment.

- optimal speed and the lane-changing speed of fleet. The capacity value changes depending on traffic flow situation and geometric factors.
- (3) The drop of merge capacity is caused by driver's lane-changing maneuvers that frequently lead to vehicle variable speed and thereby time headway was extended relative to the minimum time headway that corresponds to the maximum capacity; the reduction of capacity is calculated by accumulative loss of time headway.
  - (4) Results compared with real data and microsimulation of on-ramp merging section show that this model produces reliable and accurate results, and the results meet the sensitivity analysis.
  - (5) Our studies results suggest that a big gap of speed between mainline and ramp vehicle should be reduced and the introduction of an accelerating lane long enough can remit this problem and improve the capacity of the on-ramp merging section.
  - (6) To improve the ramp control strategies can create smaller capacity drops, when compared with no control case that is unable to prevent excessive vehicle into the mainline by the ramp.
  - (7) Introducing the design of lane number balance (the number of mainline lanes on the downstream of ramp is equal to the sum of ramp lanes and mainline lanes on the upstream) can increasing the capacity of on-ramp merging section by decreasing driver's lane-changing maneuvers.
  - (8) This study about on-ramp merging section capacity of urban expressway is important for transportation planning, design, and operations.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This research is supported by the Central South University Science Fund Project, Hunan province, China, the Science and Technology Plans Project of Hunan province, Hunan province, China, and the Open Fund of Hunan Province University Key Laboratory of Bridge Engineering (Changsha University of Science and Technology), Hunan province, China.

## References

- [1] M. J. Cassidy and J. Rudjanakanoknad, "Increasing the capacity of an isolated merge by metering its on-ramp," *Transportation Research B*, vol. 39, no. 10, pp. 896–913, 2005.
- [2] J. A. Laval and C. F. Daganzo, "Lane-changing in traffic streams," *Transportation Research B*, vol. 40, no. 3, pp. 251–264, 2006.
- [3] K. Chung, J. Rudjanakanoknad, and M. J. Cassidy, "Relation between traffic density and capacity drop at three freeway bottlenecks," *Transportation Research B*, vol. 41, no. 1, pp. 82–95, 2007.
- [4] S. Oh and H. Yeo, "Microscopic analysis on the causal factors of capacity drop in highway merging sections," in *Proceedings of the Transportation Research Board 91st Annual Meeting*, Washington, DC, USA, 2012.
- [5] A. Duret, J. Bouffier, and C. Buisson, "Onset of congestion from low-speed merging maneuvers within free-flow traffic stream," *Transportation Research Record*, no. 2188, pp. 96–107, 2010.
- [6] B. Coifman and S. Kim, "Extended bottlenecks, the fundamental relationship, and capacity drop on freeways," *Transportation Research A*, vol. 45, no. 9, pp. 980–991, 2011.
- [7] D. R. Drew, "Gap acceptance in the freeway merging process," *Highway Research Record* 208, 1967.
- [8] S. Westphal, "Capacity of freeway on-ramps on German motorways," in *Proceeding of the 2nd International Symposium on Highway Capacity*, pp. 707–715, Sydney, Australia, 1994.
- [9] P. Lertworawanich and L. Elefteriadou, "A methodology for estimating capacity at ramp weaves based on gap acceptance and linear optimization," *Transportation Research B*, vol. 37, no. 5, pp. 459–483, 2003.
- [10] S. G. Kim and Y. Son, "Development of new merge capacity model and effects of ramp flow on the merge capacity," in *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*, pp. 1–19, Washington, DC, USA, 2003.
- [11] L. Leclercq, J. A. Laval, and N. Chiabaut, "Capacity drops at merges: an endogenous model," *Transportation Research B*, vol. 45, no. 9, pp. 1302–1313, 2011.
- [12] M. R. Lorenz and L. Elefteriadou, "Defining freeway capacity as function of breakdown probability," *Transportation Research Record*, no. 1776, pp. 43–51, 2007.
- [13] W. Brilon, J. Geistefeldt, and H. Zurlinden, "Implementing the concept of reliability for highway capacity analysis," *Transportation Research Record*, no. 2027, pp. 1–8, 2007.
- [14] C. Zhao, W. Wang, and W. Li, "On-ramp capacity under different flow rate of major stream," *Journal of Highway and Transportation Research and Development*, vol. 22, no. 2, pp. 82–85, 2005.
- [15] Y. Ci, L. Wu, Y. Pei, and X. Ling, "Gap acceptance capacity model for on-ramp junction of urban freeway," *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 4, pp. 116–119, 2009.

## Research Article

# Research and Application of the Beijing Road Traffic Prediction System

Ruimin Li,<sup>1</sup> Hongliang Ma,<sup>1</sup> Huapu Lu,<sup>1</sup> and Min Guo<sup>2</sup>

<sup>1</sup> Department of Civil Engineering, Tsinghua University, Beijing 100084, China

<sup>2</sup> Beijing Traffic Management Bureau, Beijing 100037, China

Correspondence should be addressed to Ruimin Li; lrmin@tsinghua.edu.cn

Received 9 November 2013; Accepted 31 December 2013; Published 18 February 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Ruimin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As an important part of the urban Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS), short-term road traffic prediction system has received special attention in recent decades. The success of ATMS and ATIS technology deployment is heavily dependent on the availability of timely and accurate estimation or prediction of prevailing and emerging traffic conditions. We studied a real-time road traffic prediction system developed for Beijing based on various traffic detection systems. The logical architecture of the system was presented, including raw data level, data processing and calculation level, and application level. Four key function servers were introduced, namely, the database server, calculation server, Geographic Information System (GIS) server, and web application server. The functions, function modules, and the data flow of the proposed traffic prediction system were analyzed, and subsequently prediction models used in this system are described. Finally, the prediction performance of the system in practice was analyzed. The application of the system in Beijing indicated that the proposed and developed system was feasible, robust, and reliable in practice.

## 1. Introduction

Along with ever-increasing motorization in China, urban road traffic systems are facing serious congestion issues, especially in the larger cities. The development of Intelligent Transportation Systems (ITS), in particular Advanced Traffic Management System (ATMS) and Advanced Traveler Information System (ATIS), plays an increasingly important role in urban traffic management. They provide various levels of traffic information and trip advisory to system users, including many ITS information service providers, enabling travelers to make appropriate and informed travel decisions. The success of ATMS and ATIS technology deployment is heavily dependent on timely and accurate estimates of the prevailing and emerging traffic conditions. To implement ATMS and ATIS to meet various traffic control, management, and operation objectives, it is necessary to develop a road traffic prediction system that utilizes advanced traffic prediction models to analyze data, especially real-time traffic data from different sources, to estimate and predict traffic conditions.

In the past few years, real-time traffic prediction systems have been studied and developed in certain cities and regions [1, 2], based on simulation or the real-time traffic detection data.

The Traffic Estimation and Prediction system (TrEPS) developed in a dynamic traffic assignment (DTA) research project initiated by the US Federal Highway Administration (FHWA) is a typical traffic prediction system based on simulation. The system is expected to be capable of estimating and predicting traffic information for real-time traffic management and control purposes to meet the information needs in the ITS context [3, 4].

Together with IBM, the Singapore Land Transport Authority (LTA) ran a pilot project from December 2006 to April 2007, with a traffic prediction tool based on historical traffic data and real-time feeds with traffic flow conditions from several sources, to predict the levels of congestion up to an hour in advance. The pilot results showed overall prediction results with above 85% accuracy. Furthermore,

when more data was available at peak hours, average accuracy reached 90% [5].

The CAPITALS project was initiated in five European cities (Brussels, Berlin, Paris, Madrid, and Rome) by using and improving existing data resources to establish a platform for information and traffic management services for administration and travelers. A traffic prediction tool was tested and the harmonisation of traffic information in Paris was completed. The five cities above extended their information platforms towards integrated mobility service platforms, in which the prediction tools were developed in Paris, Madrid, and Berlin. In Madrid, estimation of travel times on the M30 motorway ring road was based on a collection of real-time traffic information from the network through detectors and TV cameras and a short-term prediction for congestion analysis. This information was processed in the M30 Traffic Control Centre and communicated via Variable Message Sign (VMS) panels to travelers [6].

As a key element of the Government's Transport 2010 Ten-Year Plan for developing and modernizing the transport system, England's National Traffic Control Centre has gathered real-time information from across the motorway network, improving driving conditions for road users by keeping them better informed and making journey times more reliable. From their website, users obtain the prediction information through the traffic forecaster [7].

The BAYERN ONLINE project launched by the Bavarian State Government in Germany developed the BayernInfo website [8], with one of its main functions providing short-term, mid-term, and long-term traffic prediction for travelers by using a traffic model called "ASDA-FOTO" [9]. Short-term prediction depends on real-time traffic, mid-term prediction depends on traffic events, and long-term prediction depends on traffic demand forecasts. For roads without detectors, the so-called assignment-based methods are applied.

Traffic prediction systems are also under research or construction for some Interstate Highways in America, a case in point being the I-4 Interstate Highway in Orlando, Florida [10]. In addition, most of the developments that have been conducted to date have been carried out in developed countries. In the last decade, many studies have been conducted on short-term traffic flow prediction models and system research in China [11–13], but no practical system has been implemented successfully in the literature to assist real-time traffic operation in cities or highways in China.

To improve traffic management efficiency, the Beijing Traffic Management Bureau (BTMB) launched several ITS systems, including the Beijing Road Traffic Prediction System (BRTPS). In this study we analyzed the development and performance of BRTPS. The system architecture was presented and analyzed in the second section, which is followed with the main functions of BRTPS in the third section. Three key prediction models used in the BRTPS were introduced in the fourth section, as well as the performance analysis in the fifth section. The final section gives a brief conclusion.

## 2. System Architecture

*2.1. Logical Architecture.* According to system requirements and existing devices and data resources, the logical architecture of the system is shown in Figure 1. The three-level logical architecture includes the following three levels.

*2.1.1. Data Resource Level.* The data resource level provides the BRTPS system with different data from various existing urban traffic detection systems in Beijing, including the loop detector of the traffic signal control system (covering about three hundred intersections within the second ring expressway), travel time detection system (covering 139 intersections within the fifth ring expressway with vehicle number plate recognition video), microwave traffic flow detection system (covering all expressways in Beijing, with a distance of about 300–800 m), probe vehicle detection system (about 20,000 taxi vehicles in Beijing), traffic accident reporting system from the Beijing Traffic Control Center, and other data resources.

*2.1.2. Data Processing and Prediction Level.* The data processing and prediction level is the core of the BRTPS. It is composed of the following parts.

Data processing module, which provides real-time reliable data for the integrated database via cleaning, coding, and preparation of different data from different sources.

Integrated database, which stores and processes data required by the system, including historical data, real-time processed detection data, prediction data, and statistical analysis results.

Model library, which stores various traffic flow prediction models, traffic accident duration time prediction models, capacity models of intersections and road segments, and analysis models.

Knowledge base, which stores the temporal-spatial relationships produced by traffic flow pattern recognition models and provides basic parameter configuration for the prediction models.

GIS platform, which displays all necessary spatial data and spatial attributes of the system.

The main products of the data processing and prediction level are the predicted values of various traffic flow parameters at different time intervals.

*2.1.3. Application Level.* The application level is composed of certain application systems supported by the BRTPS, including Personalized Trip Planning and Guiding System and the traffic management system of the traffic control center, and information service providers.

*2.2. Physical Architecture.* Based on Microsoft .Net Remoting technique, the distributed physical architecture of the system is presented and shown in Figure 2.

The main components of the physical architecture are the four servers, which perform the core functions of the system.

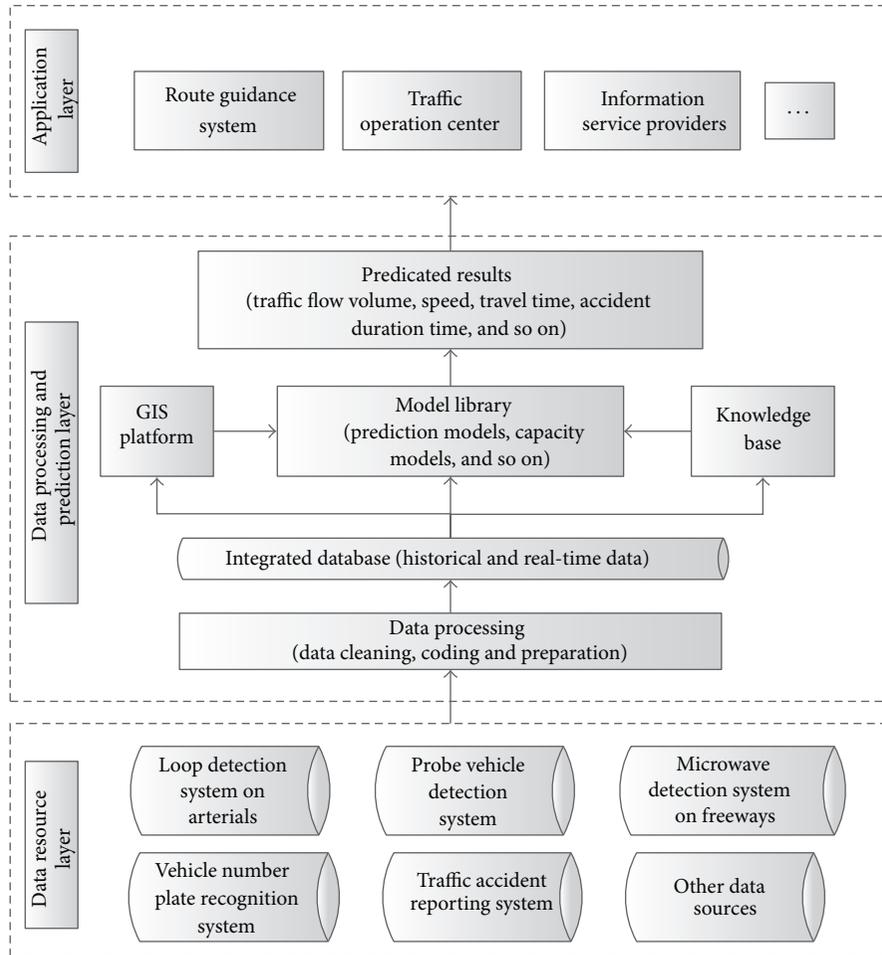


FIGURE 1: Logical architecture of BRTPS.

2.2.1. *Database Server.* The database server keeps the integrated database running, with the following main functions: (1) obtaining raw data from the existing data center, performing data processing, which transforms the raw data into standardized basic data required by the system, and storing the basic data into the integrated database; (2) storing all necessary basic data and results of traffic flow conditions required by the system; and (3) responding to the requests of reading, writing, and updating traffic flow conditions data from the other three servers.

2.2.2. *Calculation Server.* The calculation server performs various prediction models used in the system, with the following main functions: (1) obtaining basic data from the database server, calculating traffic flow prediction, road network level of service evaluation, congestion evaluation, incident warning, and temporal-spatial influence analysis based on those data, and then sending the prediction results to the database server; (2) responding to control requests from the web application server by performing requested configuration and thus changing the calculation logic; and (3) responding to the calculation requests from the web application server by performing requested calculations and then sending the results to the web application server.

2.2.3. *GIS Server.* The main functions of the GIS server include (1) storing urban road network geographical data required by the system; (2) responding to requests of the web application server by analyzing requirements for GIS data and traffic flow data, obtaining the latter from the database server and combining them with GIS data to obtain visualization information, and then sending the visualization information to the web application server; and (3) responding to requests to modify GIS information from the web application server.

2.2.4. *Web Application Server.* The web application server deals with requests from the other terminals on the network by interpreting requests into requests on GIS data, traffic flow data, and calculation, sending the requests to the other three servers accordingly, and providing user web information based on the information returned from the other servers.

This system provides service via its graphical user interface: system users visit the web application server from their terminals and send requests to the web application server from the browser, which will be analyzed and interpreted by the web application server and sent to the other three servers; these servers will then return the results to the web application server for final processing and displaying on the website for the users.

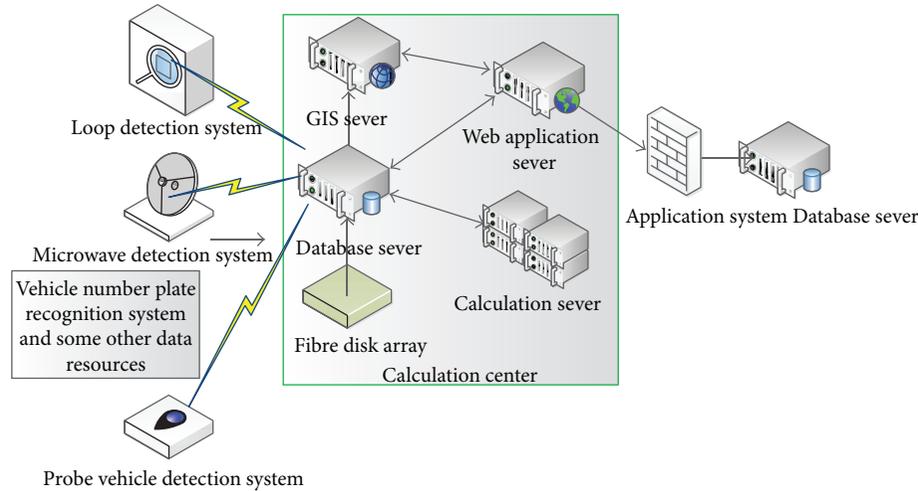


FIGURE 2: Physical Architecture of the system.

This system also provides service by delivering results for other application systems: based on the requirements of these systems, this system will send the prediction results to these systems at the same time as storing the results in its own integrated database, or other systems obtain the prediction results regularly from the integrated database of this system before performing their own processing and application according to their own needs.

### 3. System Functions

The system mainly consists of the following functions.

*Traffic Prediction under Normal Conditions and Prediction Model Update.* Based on the integrated database, the system will predict traffic flow conditions with different intervals using various traffic prediction models. Every five minutes, traffic flow parameters, including flow volume, speed, occupancy, and travel time, are predicted with time intervals of 5 mins, 15 mins, 30 mins, 1 h, and 2 h.

Traffic flow prediction models are updated online in accordance with the operation of the system. The correction factors in various prediction models, such as weight factors in the combined prediction model, are continuously adjusted according to the prediction performance or the traffic condition changes to improve prediction accuracy and the model's adaptability to various traffic conditions.

*Temporal-Spatial Influence Analysis and Prediction of Traffic Accidents.* Based on real-time detected traffic flow data and accident information from the traffic accident reporting system, this system analyzes the temporal-spatial influence of traffic accidents in the Beijing road network. It provides predicted duration time and influence scope of an accident for urban road traffic management administrators.

*Traffic Flow Condition Analysis and Evaluation.* The system also analyzes and evaluates urban road traffic conditions at

the road section, intersection, and region level by adjusting traffic condition evaluation factors and assessing the transport level of service. It also analyzes the detected and predicted data to evaluate the level of traffic congestion.

*Urban Road Traffic Changing Trend Analysis.* The system can identify the traffic flow changing trend both temporally and spatially, with the immense amount of traffic flow data stored in the system's database. It analyzes the characteristics and trends of traffic flow in different regions, intersections, and sections and the correlation of the traffic flow between them, to provide support for urban road traffic management administrators.

*Traffic Information Service.* The system can generate traffic flow condition assessment and prediction information, which may be provided for other urban road traffic management systems, organizations, or individuals who have an interest, for example, information service providers. Additionally, it can also disseminate prediction information to public travelers through the VMS or the internet.

### 4. Key Prediction Models

To develop a practical system that can be deployed in the BBTM traffic control center, we presented and modified several models, including the traffic flow parameter correlation model, the capacity calculation model for expressways, urban arterials and intersections, the traffic flow parameter prediction models under normal traffic flow conditions, the Automatic Incident Detection (AID) model, and the accident temporal-spatial influence analysis model [14]. Here we introduce two traffic flow parameter prediction models under normal traffic flow conditions and the accident duration time prediction model.

*4.1. Combined Traffic Flow Prediction Model.* During the development of the Beijing Road Traffic Prediction System,

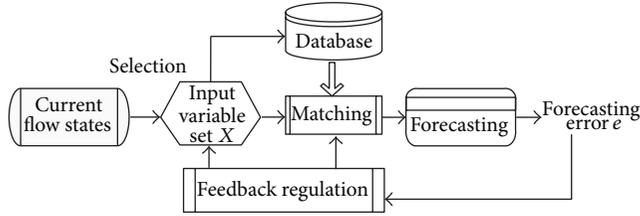


FIGURE 3: Forecasting frame based on NPR.

to find the most suitable prediction model for Beijing's traffic flow conditions, various short-term traffic flow prediction models were proposed for detected and nondetected roads, including the combined traffic flow prediction model [15], the nonparametric regression model [16], and the combined neural network prediction model [17]. The former two models were applied in the system according to the consideration of computation efficiency and prediction accuracy.

The combined prediction model for the BRTPS was considered with the composition of the Discrete Fourier transform model (DFT), Autoregressive model (AR), and Neighborhood Regression model (NR). For convenience, we denoted DFT-AR-NR as the DAN model [15]. Traffic prediction for road sections was not only associated with its historical and recent data of the road section of interest but also with data from adjacent sections. Therefore, a basic form of the DAN model can be represented as [15]

$$x = \alpha x^+ + \beta x^\wedge + \gamma x^*, \quad (1)$$

where  $x^+$ ,  $x^\wedge$ , and  $x^*$  denote the prediction results of the three submodels, respectively, and  $\alpha$ ,  $\beta$ , and  $\gamma$  are the weight coefficients of the three submodels, respectively. Adjusting the value of these weight coefficients can strengthen or weaken the role for any of the submodels. The DAN model was mainly used for detected road segments.

**4.2. Nonparametric Regression Model.** The short-term traffic flow forecasting frame based on nonparametric regression is shown in Figure 3 [16].

The whole system process is as follows.

- (1) The system input variable sets were determined by the selection algorithm of current flow states.
- (2) The input variable set  $X$  was matched among the flow states stored in database to find  $K$  optimal matching states. If forecasting time was ample, the linear matching algorithm was the best choice; otherwise we resorted to nonlinear matching algorithm and complex data structure, for example, binary tree and R tree.
- (3) The successfully matching states  $Y$  were averaged to obtain the forecasting values.
- (4) The forecasting error  $e$  was put into the feedback regulation module to adjust the input variable set and matching algorithm.

The nonparametric regression based model was mainly used for nondetected road segments.

**4.3. Traffic Accident Duration Time Prediction Model.** For traffic accident duration time prediction, a model based on the algorithm of decision tree, Classification and Regression Tree (CART), was presented and applied [18]. The model was developed based on accident records extracted from the accident reporting system of the Beijing Traffic Management Bureau. When an accident occurred, this model will be used to predict the duration of the accident.

## 5. System Deployment and Performance Analysis

**5.1. System Deployment.** Based on the above models and various data resources, the Beijing Road Traffic Prediction System was developed in the following environment: database system: ORACLE 10g database, web server: IIS6, and WebGIS developing and operating system: ArcGIS Server 9.0 from ESRI. The client uses Windows 98 OS or above and web browser IE6.0 or above.

Before the 2008 Olympic Games, the 1.0 vision of BRTPS mainly covered 14 detected expressways and arterial streets within the second ring expressway and was deployed in the traffic control center of BBTM for normal traffic conditions.

In 2011, this system was updated to cover all expressways and arterial streets within the fifth ring expressway, for normal and event traffic conditions. Figure 4 shows the BRTPS interface.

The data used in the system mainly comes from the expressway traffic flow detection system (microwave detectors), travel time detection system based on vehicle number plate recognition, traffic signal control system detectors, and floating car system based on taxi and accident reporting system as mentioned above. It predicts traffic parameters such as flow, speed, and occupancy in 5 min, 15 min, 30 min, 1 h, and 2 h intervals.

**5.2. System Performance Analysis.** To understand the prediction performance of the practical system, prediction error analysis was carried out during November 2012.

Fifteen sites selected for the application of the DAN model included ten different expressways in Beijing. Most sites are very congested during morning and evening peak hours. Ten days were selected as test days for all fifteen sites, namely, November 12–16, 2012, and November 26–30, 2012. From 7:00 to 13:00 and from 14:00 to 19:00 every day, we selected the detected data and the predicted data hourly. The predicted data included the predicted value of traffic flow, speed, and occupancy in 5 min, 15 min, and 30 min intervals. We mainly analyzed the error performance of speed prediction, which was the most precise among the three traffic flow parameters of volume, speed, and occupancy.

For analysis of system prediction performance, mean absolute percentage error (MAPE) and mean absolute error (MAE) were selected and employed to reflect the accuracy of the predictor.

TABLE 1: MAPE of speed prediction of the largest error.

Site ID	Prediction interval (mins)	Hour	Speed prediction value	Detected speed value	Traffic flow volume	Occupancy value (%)	Prediction error
13	30	17	45.91	9.83	1296	83.33	367.04%
13	5	18	39.16	8	1620	66.67	389.5%
10	15	17	63.77	12.67	4764	47.44	403.35%
3	5	14	56.95	11.17	3348	71.17	409.85%
7	15	9	64.73	12.67	1440	52.67	410.89%
10	5	18	59.45	11.5	3762	49.83	416.96%
1	15	10	50.32	9.08	2742	65.5	454.17%
3	15	18	54.71	9.17	2748	70.67	496.62%
7	5	19	58.79	9.33	438	63.33	530.12%
13	30	17	61.21	9.08	2040	66.67	574.12%



FIGURE 4: System interface.

MAPE and MAE are defined as follows:

$$\text{MAPE} = \frac{\sum_{t=0}^{N-1} \left( \frac{\text{abs} [V(t+1) - \widehat{V}(t+1)]}{V(t+1)} \right)}{N},$$

$$\text{MAE} = \frac{\sum_{t=0}^{N-1} \left( \text{abs} [V(t+1) - \widehat{V}(t+1)] \right)}{N}, \quad (2)$$

where  $V(t+1)$  is the observed traffic flow speed for the time interval  $t+1$ ,  $\widehat{V}(t+1)$  is the predicted traffic flow speed for the time interval  $t+1$ , and  $N$  is the number of intervals for prediction.

The MAPE of speed prediction for different intervals of the fifteen sites over ten days is shown in Figure 5.

From Figure 5, the average MAPE of speed prediction over the ten days increased slowly with increasing prediction interval, specifically by 14.5% for 5 min interval, 16.4% for 15 min, and 16.8% for 30 min. Eleven sites had MAPE speed prediction at the 5 min interval below 20%, eleven sites at the 15 min interval, and ten sites at the 30 min interval. Thus, speed prediction performance of most selected sites was satisfactory.

The MAPE and MAE of speed prediction at different hours are shown in Figures 6 and 7, respectively. There were no apparent differences in performance in different hours, except for the afternoon peak hours of 17:00 and 18:00. Both MAPE and MAE during afternoon peak hours were larger than that during other hours. The larger errors during the

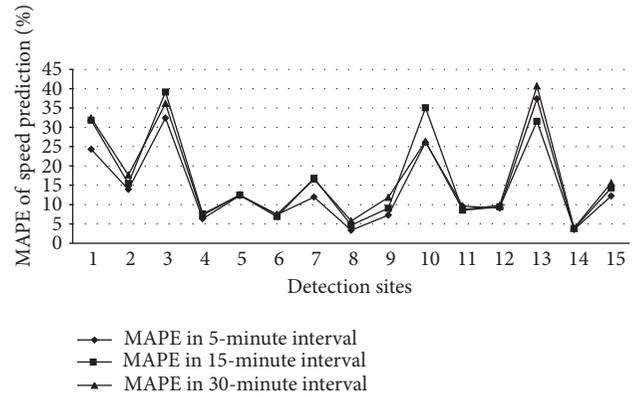


FIGURE 5: MAPE of speed prediction in different intervals of fifteen sites in ten days.

afternoon peak hours indicated that the models deployed in the system may need to be improved for congestion conditions in the future or for some road segments.

The MAPE of speed prediction of the selected sites shows that the accuracy of BRTPS is similar with some other systems, for example, the traffic prediction tool developed by IBM Research for Singapore, in which the overall prediction results were well above the target accuracy of 85 percent [5].

10 speed prediction values with largest prediction error in 5400 data are listed in Table 1, in which the same hour for the same site ID indicates that the hour is in different days.

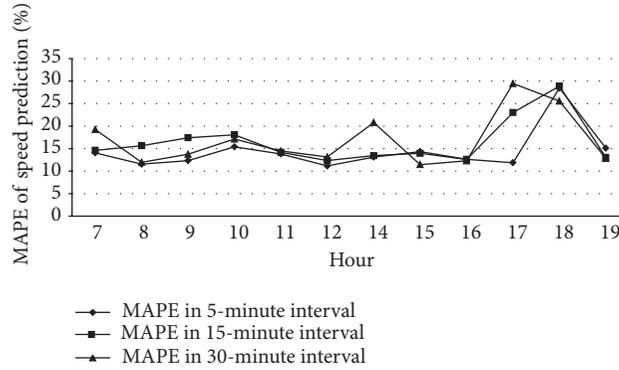


FIGURE 6: MAPE of speed prediction in different hours.

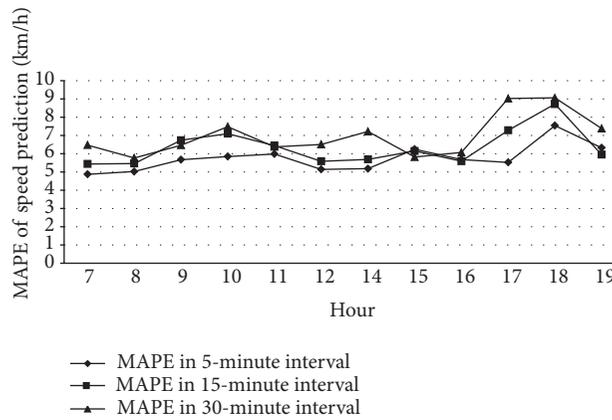


FIGURE 7: MAE of speed prediction in different hours.

From Table 1, we can see that the ten cases with large prediction error almost were undersaturated condition, as shown with the speed and occupancy values. These large speed prediction errors may resulted from two reasons. The first is that the prediction model cannot deal with the traffic condition changing from undersaturated to oversaturated. For example, at 16:45, the traffic condition is free flow and at 17:00 the traffic flow suddenly becomes congested; the combined prediction cannot suit for the changing well. On the other hand, the traffic flow condition at 17:00 may be caused by an event, for example, an accident, and in the current application system did not consider the effect of special event in the prediction model before the event occurred.

### 6. Conclusions

Real-time traffic prediction systems are one of the foundations of ATMS and ATIS. We studied the logic structure, physical structure, and main functions of the Beijing Road Traffic Prediction System deployed in the control center of BTMB. The key prediction models and the online prediction performance were also introduced. Performance analysis indicated that the system satisfied prediction accuracy most

of the time for expressways. As discussed, however, during the application period the current system may sometimes produce larger prediction errors, especially during the transition period from free-flow to congested traffic or under congestion conditions. Future prediction accuracy may be improved by refining the developed model based on detected data or by integrating other prediction models based on real-time dynamic traffic assignment.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This work was supported by Committee of Beijing Science and Technology (Grant no. D070206014007) and National Key Technology R&D Program (Grant no. 2009BAG13A06).

### References

[1] C. Alecsandru and S. Ishak, "Hybrid model-based and memory-based traffic prediction system," *Transportation Research Record*, no. 1879, pp. 59–70, 2004.

- [2] B. S. Kerner, H. Rehborn, M. Aleksić, and A. Haug, "Traffic prediction systems in vehicles," in *Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems*, pp. 251–256, Vienna, Austria, September 2005.
- [3] H. Lieu, *A Roadmap for the Research, Development and Deployment of Traffic Estimation and Prediction Systems for Real-Time and Off-Line Applications (TrEPS, TrEPS-P)*, Travel Management Team Office of Operations RD&T, Federal Highway Administration, 2003.
- [4] M. Ben-Akiva, M. Bierlaire, H. Koutsopoulos, and R. Mishalani, "DynaMIT: a simulation-based system for traffic prediction," in *Proceedings of the DACCORD Short Term Forecasting Workshop*, Delft, The Netherlands, February 1998.
- [5] "IBM and Singapore's Land Transport Authority Pilot Innovative Traffic Prediction Tool," IBM Press Release, 2007, <http://www-03.ibm.com/press/us/en/pressrelease/21971.wss>.
- [6] CAPITALS PLUS TR 4029, 2013, <http://cordis.europa.eu/telematics/tap-transport/research/projects/capitalsplus.html>.
- [7] Highway Agency, 2013, <http://www.highways.gov.uk/traffic-information/>.
- [8] BayernInfo, 2012, <http://bayerninfo.de/en>.
- [9] B. S. Kerner, H. Rehborn, M. Aleksic, and A. Haug, "Recognition and tracking of spatial-temporal congested traffic patterns on freeways," *Transportation Research Part C*, vol. 12, no. 5, pp. 369–400, 2004.
- [10] S. Ishak and H. Al-Deek, "Statistical Evaluation of Interstate 4 Traffic Prediction System," *Transportation Research Record*, no. 1856, pp. 16–24, 2003.
- [11] F. You, W. Yu, and H. Rong, "Research on the traffic flow prediction system framework for ITS based on organic computing," in *Proceedings of the 10th International Conference of Chinese Transportation Professionals (ICCTP '10)*, pp. 2180–2191, August 2010.
- [12] L. Yong and S. Houbing, "DynaCHINA: a specially-built real-time traffic prediction system for China," in *Proceedings of the TRB 86th Annual Meeting Compendium of Papers CD-ROM*, Washington, DC, USA, 2007.
- [13] Y. Li, X. Hao, and L. Liu, "Short-term traffic flow forecasting based on the improved non-parametric regression," in *Proceedings of the 2nd International Conference on Transportation Information and Safety (ICTIS '13)*, pp. 914–920, 2013.
- [14] L. Huapu, S. Yagang, and G. Min, *Urban Road Mixed Traffic Flow Analysis Model and Method*, China Railway Publishing House, Beijing, China, 2009.
- [15] S. Dong, R. Li, L. G. Sun, T. H. Chang, and H. Lu, "Short-term traffic forecast system of Beijing," *Transportation Research Record*, no. 2193, pp. 116–123, 2010.
- [16] X.-L. Zhang and H.-P. Lu, "The simulation research of non-parametric regression for short-term traffic flow forecasting," in *Proceedings of the International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '09)*, vol. 3, pp. 626–629, Zhangjiajie, China, April 2009.
- [17] R. Li and H. Lu, "Combined neural network approach for short-term urban freeway traffic flow prediction," in *Proceedings of the 6th International Symposium on Neural Network (ISNN '09)*, vol. 5553 of *Lecture Notes in Computer Science*, part 3, pp. 1017–1025, May 2009.
- [18] R. Li, X. Zhao, X. Yu, J. Li, N. Cheng, and J. Zhang, "Incident duration model on urban freeways using three different algorithms of decision tree," in *Proceedings of the International Conference on Intelligent Computation Technology and Automation (ICICTA '10)*, vol. 2, pp. 526–528, Changsha, China, May 2010.

## Research Article

# Dynamic Network Design Problem under Demand Uncertainty: An Adjustable Robust Optimization Approach

Hua Sun,<sup>1</sup> Ziyu Gao,<sup>2</sup> and Fangxia Zhao<sup>1</sup>

<sup>1</sup> MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup> Institute of System Science, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Hua Sun; sunhua19800601@163.com

Received 31 October 2013; Accepted 18 December 2013; Published 17 February 2014

Academic Editor: Huimin Niu

Copyright © 2014 Hua Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper develops an adjustable robust optimization approach for a network design problem explicitly incorporating traffic dynamics and demand uncertainty. In particular, a cell transmission model based network design problem of linear programming type is considered to describe dynamic traffic flows, and a polyhedral uncertainty set is used to characterize the demand uncertainty. The major contribution of this paper is to formulate such an adjustable robust network design problem as a tractable linear programming model and justify the model which is less conservative by comparing its solution performance with the robust solution from the usual robust model. The numerical results using one network from the literature demonstrate the modeling advantage of the adjustable robust optimization and provided strategic managerial insights for enacting capacity expansion policies under demand uncertainty.

## 1. Introduction

Over the past three decades, network design problem (NDP) has been pervasively studied and applied in the different fields, such as transportation, production, distribution, and communication. The different research field is corresponding to different sets of objectives, decision variables, and resource constraints, implying different behavioral and system assumptions, and possessing varying parameters and data requirements and capabilities in terms of representing the network supplies and demands [1]. NDP models have been extensively used as a decision-making tool and provide the guidelines for enacting the capacity expansion policies. In transportation, network design models in general aim at maximizing the social benefit (or minimizing the total travel costs) through implementing an optimal capacity expansion policy in the network.

However, traditional NDP models for transportation networks assume that the demand or/and capacity are deterministic. Obviously, this is unrealistic, because the real-life transportation network systems are surrounded by a number of uncertainties, both from the supply side and demand side.

Evaluation of network performance without accounting for demand uncertainty can potentially lead to biased investment decisions [2]. Therefore, recently, NDPs under uncertainty have drawn increasing attentions. Chen et al. [3] provided a state-of-the-art review of the transportation network design problem under uncertainty.

However, the vast body of the literature in the past has focused on the static NDPs [4–6]. Lin et al. [7] pointed out the three drawbacks of the static NDP models when compared with the dynamic NDP (DNDP) models. Therefore, to overcome these deficiencies, a variety of recent papers have focused on dynamic NDP models [7–11]. In dynamic NDP models, dynamic traffic assignment (DTA) models are used to model the time-dependent variation of traffic flows and travel behaviors and characterize the transportation network flow pattern. Janson [12] and Waller [13] showed that the DTA-based NDP model is more desirable than the static model. According to their formulations, DTA-based NDP models can be classified into two categories: single-level models and bilevel models [7]. The focus of this paper is on the application adjustable robust optimization (ARO) for

the single-level dynamic NDP under demand uncertainty, where the single-level structure is adopted because it can provide an easier way to manipulate affinely adjustable robust counterpart (AARC) and make adjustable robust dynamic NDP (ARDNDP) to be computationally tractable.

There is an increasing body of relevant literature on the DTA-based NDP under uncertainty [1, 2, 10, 14–18]. The common feature of these studies is that the cell transmission model (CTM) [19, 20] is adopted to model the time-varying traffic flows propagation, and the traffic demand is assigned to the network by either the dynamic system-optimal (SO) [21] or user-optimal (UO) [10] principle. The approaches of addressing uncertainty in above-mentioned studies mainly are stochastic programming (SP) and robust optimization (RO). However, these approaches suffer from deficiencies related to lack of data availability or are overly conservative, which seriously limit their applicability in traffic management and control.

In this paper, we employ the ARO approach to develop the single-level CTM-based system-optimal NDP model. Given the fact that the single-level CTM-based NDP model has a linear programming formulation, we use the ARO method [22] to formulate the ARDNDP, which overcomes the limitation of stochastic optimization and usual robust optimization [23–27]. Specifically, not only is there no such assumption that the probability distribution of uncertain demand is known but also it is less conservative than robust dynamic NDP in our model. Similar to robust optimization approach, we only need to simply specify a closed set to limit the demand uncertainty, which is readily implemented in most applications. The adjustable robust counterpart (ARC) of system-optimal DNDP (SODNDP) can be reformulated as the computationally tractable affinely adjustable robust counterpart (AARC) of SODNDP by using prescribed uncertainty set and the affine decision rule.

We emphasize the main contributions of this paper at a glance below.

We provide an ARO framework for the SO DTA-based ARDNDP. Note that we describe the ARDNDP model only for single-destination, system-optimal networks. However, this is unrealistic. How to extend the basic ARCC formulation method to the multidestination problem case is our future research direction.

An appealing feature of our AARC is that it is less conservative than the usual RC. Moreover, AARC is still a linear programming model. So, in general, it is computationally tractable and can be solved in polynomial time by a few well-known solution algorithms.

Our numerical example demonstrates the value of ARO in the context of dynamic NDP. The computation viability is demonstrated for the proposed modeling framework. The numerical analysis for the impact of investment budget bound and the demand uncertainty level on the network design solution justifies the solution robustness and flexibility.

The remaining part of this paper is structured as follows. Section 2 provides a discussion of the relevant literature. In Section 3, we describe the formulation given by [21] as a CTM-based deterministic dynamic NDP. Section 4 proposes

an affinely adjustable robust counterpart formulation of the dynamic NDP under demand uncertainty by using the affine decision rule. In Section 5, we develop a traditional robust counterpart formulation of the dynamic NDP under demand uncertainty. Numerical experiments and results analysis are elaborated in Section 6. Finally, Section 7 concludes the paper and proposes the potential future extensions.

## 2. Literature Review

Dynamic NDP is at the core of many transportation planning problems and has been extensively studied in the literature. Although a great amount of attention has been paid to the deterministic dynamic NDP in the past years, increasing efforts are focusing on incorporating uncertainty in the dynamic NDP because the researchers have realized that input data and parameters have the inherently uncertain nature and uncertain input data and parameters may have a drastic impact on the optimality and the feasibility of the solution. Therefore, various modeling techniques are applied to cope with uncertainty, which mainly include Monte-Carlo sampling approach, chance-constrained programming (CCP), a two-stage stochastic programming with recourse (SLP2), scenario-based robust optimization, and set-based robust optimization. Karoonsoontawong and Waller [15] applied Monte-Carlo bounding techniques to solve the stochastic SO and UO dynamic NDP. They showed that it is more beneficial to solve a stochastic model than a deterministic model. Waller and Ziliaskopoulos [14] formulated the CCP and SLP2 models of single-level CTM-based SO NDP under demand uncertainty. Ukkusuri and Waller [10] introduced the CTM-based single-level UO versions of CCP and SLP2 and compared them with the corresponding SO versions. Karoonsoontawong and Waller [17] proposed a SLP2 model of CTM-based bilevel NDP. Karoonsoontawong and Waller [18] further developed a SLP2 model of CTM-based bilevel combined NDP and signal setting design (SSD). The above-mentioned CTM-based NDP models were, however, developed by the Monte-Carlo sampling, CCP, or SLP2 approach and it is necessary for the modeler users to know the probability distributions of the uncertain input data and parameters in order to use these models. In fact, the distributions may be unavailable (inaccurate) in reality because we may have no (insufficient) data to calibrate the distributions. Therefore, robust optimization was introduced recently to address the limitations of CTM-based NDPs or DTA. According to Chung et al. [1], robust optimization can be roughly classified into two groups: (1) scenario-based robust optimization and (2) set-based robust optimization.

Mulvey et al. [28] developed a scenario-based RO approach for general linear programming. Karoonsoontawong and Waller [16] used this approach to propose the CTM-based bilevel NDP model with exact solution methods and developed three metaheuristics for multiorigin multi-destination large problem: simulated annealing, genetic algorithm, and random search. Karoonsoontawong and Waller [17] adopted the same approach to formulate CTM-based single-level SO and UO versions of NDP models and bilevel

NDP model and compared them. Karoonsoontawong and Waller [18] presented a bilevel scenario-based robust model for the CTM-based combined NDP and SSD. Similar to Monte-Carlo sampling, CCP, and SLP2, the scenario-based RO approach also need to know the probability of each scenario in advance. In additional, the numerous scenarios used in accurately representing the uncertainty can lead to large, computationally challenging problems and the solution obtained may be sensitive to possible uncertainty outcomes.

To overcome the limitation of Monte-Carlo sampling, CCP, SLP2, and scenario-based RO approaches, the set-based RO approach has been applied to cope with uncertain input data and parameters. Unlike Monte-Carlo sampling, CCP, SLP2, and scenario-based RO approaches, the set-based RO approach [23–27] does not require the assumption that the probability distributions of the uncertain input data and parameters are known, while the uncertain input data and parameters are assumed to belong to a bounded and convex set. Therefore, recently, the set-based RO has been applied to not only static NDP [29–31] but also CTM-based dynamic NDP or DTA. In CTM-based dynamic NDP or DTA, Chung et al. [1] applied the set-based RO approach to formulate a single-level robust NDP model under demand uncertainty, in which the demand uncertainty is assumed to belong to a box set. Yao et al. [32] used the polyhedral, box, and ellipsoid as uncertain set to develop the CTM-based system-optimal DTA (SODTA). Chung et al. [33] used the set-based RO approach to develop dynamic congestion pricing under dynamic user equilibrium. They proposed a bilevel cellular particle swarm optimization (BSCPO) algorithm to solve the corresponding robust counterpart and compared it with bilevel simulated annealing (BSA) and cutting plane-based simulated annealing (CPSA) algorithms. They showed that the BSCPO algorithm proposed outperforms these two alternative algorithms through test results. However, the robust solutions obtained in the above set-based RO studies are overly conservative. Therefore, Ben-Tal proposed an adjustable robust optimization formulation (ARO) for general linear programming that can alleviate the conservatism of robust solutions obtained through the usual RO approach. Ben-Tal et al. [34] applied the ARO methodology to solve the CTM-based SODTA under demand uncertainty. The polyhedral set is adopted as the uncertain demand set, and the affinely adjustable robust counterpart (AARC) was reformulated as linear programming by using the affine decision rule and duality theory. In this paper, we extend ARO approach to the CTM-based SO NDP model and provide guidance for making capacity expansion plans under uncertainty in context of transportation network design problem.

### 3. Deterministic Model

In this section, we present the deterministic version of the CTM-based single-level SO dynamic NDP model, which provides the basic modeling platform and functional form the ARDNDP model we will introduce in the next section. For discussion convenience, we present the notation used throughout these models (see Notations).

The CTM-based single-level SO dynamic NDP (SODNDP) model aims at minimizing the total travel cost, which is the sum of the product of the number of vehicles in each cell in each time interval and the corresponding travel cost. To penalize unmet demand by the end of time horizon  $T$ , the travel cost in cell  $i$  at time  $t$ ,  $c_i^t$ , is set as follows:

$$c_i^t = \begin{cases} 1 & i \in C \setminus C_s, t \neq T \\ M & i \in C \setminus C_s, t = T, \end{cases} \quad (1)$$

where  $M$  is a sufficiently large positive number to represent the penalty cost, which can be interpreted as the cost of a vehicle that cannot arrive at the destination by the end of time horizon  $T$ . Use of the penalty cost has the effect of minimizing the number of vehicles staying in the network by the end of the time horizon  $T$ . By assuming SO principle and the linear relationship between investment and capacity increase and using the notation in Notations, the deterministic CTM-based single-level SODNDP can be formulated as the following linear programming [9]:

SODNDP:

$$\min_{x,y,b} \sum_{t \in \mathfrak{T}} \sum_{i \in C \setminus C_s} c_i^t x_i^t \quad (2)$$

$$\text{subject to } x_i^t - x_i^{t-1} - \sum_{k \in C} a_{ki} y_{ki}^{t-1} + \sum_{j \in C} a_{ij} y_{ij}^{t-1} = d_i^{t-1}, \quad \forall i \in C_R, t \in \mathfrak{T}, \quad (3)$$

$$x_i^t - x_i^{t-1} - \sum_{k \in C} a_{ki} y_{ki}^{t-1} + \sum_{j \in C} a_{ij} y_{ij}^{t-1} = 0, \quad \forall i \in C \setminus C_R \cup C_s, t \in \mathfrak{T}, \quad (4)$$

$$\sum_{j \in C} a_{ij} y_{ij}^t \leq x_i^t, \quad \forall i \in C \setminus C_s, t \in \mathfrak{T}, \quad (5)$$

$$\sum_{k \in C} a_{ki} y_{ki}^t \leq \delta_i^t (N_i^t + \chi_i b_i - x_i^t), \quad \forall i \in C \setminus C_R \cup C_s, t \in \mathfrak{T}, \quad (6)$$

$$\sum_{k \in C} a_{ki} y_{ki}^t \leq Q_i^t + \phi_i b_i, \quad \forall i \in C \setminus C_R \cup C_s, t \in \mathfrak{T}, \quad (7)$$

$$\sum_{j \in C} a_{ij} y_{ij}^t \leq Q_i^t + \phi_i b_i, \quad \forall i \in C \setminus C_s, t \in \mathfrak{T}, \quad (8)$$

$$x_i^0 = \hat{x}_i, \quad \forall i \in C \setminus C_s, \quad (9)$$

$$y_{ij}^0 = 0, \quad \forall (i, j) \in C \times C, \quad (10)$$

$$x_i^t \geq 0, \quad \forall i \in C \setminus C_s, t \in \mathfrak{T}, \quad (11)$$

$$y_{ij}^t \geq 0, \quad \forall (i, j) \in C \times C, \quad t \in \mathfrak{T}, \quad (12)$$

$$b_i \geq 0, \quad \forall i \in C \setminus C_S, \quad (13)$$

$$\sum_{i \in C \setminus C_S} b_i \leq B. \quad (14)$$

The objective function of SONDP presents the total travel cost, which provides an optimistic estimate or lower bound of total cost as it simplifies the original CTM model by Daganzo [19, 20] and allows vehicle holding [35]. Both constraints (3) and (4) are the flow conservation constraints in cell  $i$  at time interval  $t$ . Because only the source cells generate demand, the right-hand side of (3) is set as  $d_i^{t-1}$  and the right-hand side of (4) is equal to 0. Constraint (5) bounds the total outflow rate of a cell by its current occupancy. Constraint (6) represents that the total inflow rate of a cell is bounded by its remaining capacity. Constraints (7) and (8) ensure that the total inflow and outflow rate of a cell are limited by the expandable inflow and outflow capacity, respectively. The remaining constraints, from (9) to (13), report the initial conditions and nonnegativity conditions. Constraint (14) guarantees that the total cost of expansion is within the available budget.

#### 4. Affinely Adjustable Robust Formulation

In this section, we firstly develop the adjustable robust counterpart (ARC) of the SODNDP model, which incorporates the demand uncertainty into a linear programming problem via the ARO approach. However, the formulation of ARC is intractable since it is a semi-infinite program. Then, we reformulate the ARC of SODNDP as a tractable linear programming by using the affine decision rules proposed by Ben-Tal et al. [22]. In the SODNDP, constraint (3) is the only set of constraints related to demand generation. Constraint (3) can be reformulated as the following inequality constraint [1, 36]:

$$x_i^t - x_i^{t-1} - \sum_{k \in C} a_{ki} y_{ki}^{t-1} + \sum_{j \in C} a_{ij} y_{ij}^{t-1} \geq d_i^{t-1}, \quad (15)$$

$$\forall i \in C_R, \quad t \in \mathfrak{T}.$$

We can easily identify the optimal solution to the problem without constraint (15) which is zero for all  $x$  and  $y$ . The introduction of constraint (15) makes zero no longer the optimal solution to the problem. Thus, constraint (15) is always binding at the optimal solution and it can replace constraint (3). Note that the vehicle holding phenomenon can also happen in this model since constraint (15) is always binding and two equations are equivalent in this model. Similarly, constraint (4) can be replaced by the following inequality constraint:

$$x_i^t - x_i^{t-1} - \sum_{k \in C} a_{ki} y_{ki}^{t-1} + \sum_{j \in C} a_{ij} y_{ij}^{t-1} \geq 0, \quad (16)$$

$$\forall i \in C \setminus C_R \cup C_S, \quad t \in \mathfrak{T}.$$

Next, we assume that uncertain demand  $d_i^{t-1}$  belongs to a prescribed uncertainty set. A box uncertainty set is used to

bound demand uncertainty in Chung et al. [1]. However, Ben-Tal et al. [34] pointed out that the box uncertainty set is overly conservative and adopted the polyhedral set as uncertain demand set. Similar to Ben-Tal et al. [34], we choose the following polyhedral set as the uncertainty set:

$$d_i^t \in U = \left\{ d_i^t : (1 - \theta) \bar{d}_i^t \leq d_i^t \leq (1 + \theta) \bar{d}_i^t, \right. \\ \left. \forall i \in C_R, \quad t \in \mathfrak{T}, \quad \sum_{t \in \mathfrak{T}} d_i^t \leq D_i, \quad \forall i \in C_R \right\}, \quad (17)$$

where  $\bar{d}_i^t$  is the nominal demand and  $\theta$  is the demand uncertainty level. Next, we use the adjustable robust optimization methods to formulate the adjustable robust counterpart of SODNDP. Then, we have

SODNDP-ARC:

$$\min_b \max_d \min_{x, y} \sum_{t \in \mathfrak{T}} \sum_{i \in C \setminus C_S} c_i^t x_i^t, \\ \text{subject to} \quad x_i^t - x_i^{t-1} - \sum_{k \in C} a_{ki} y_{ki}^{t-1} \\ + \sum_{j \in C} a_{ij} y_{ij}^{t-1} \geq d_i^{t-1}, \quad (18)$$

$$\forall i \in C_R, \quad t \in \mathfrak{T}, \quad \forall d_i^t \in U,$$

$$x_i^t - x_i^{t-1} - \sum_{k \in C} a_{ki} y_{ki}^{t-1} \\ + \sum_{j \in C} a_{ij} y_{ij}^{t-1} \geq 0$$

$$\forall i \in C \setminus C_R \cup C_S, \quad t \in \mathfrak{T},$$

$$\sum_{j \in C} a_{ij} y_{ij}^t \leq x_i^t$$

$$\forall i \in C \setminus C_S, \quad t \in \mathfrak{T},$$

$$\sum_{k \in C} a_{ki} y_{ki}^t \leq \delta_i^t (N_i^t + \chi_i b_i - x_i^t)$$

$$\forall i \in C \setminus C_R \cup C_S, \quad t \in \mathfrak{T},$$

$$\sum_{k \in C} a_{ki} y_{ki}^t \leq Q_i^t + \phi_i b_i$$

$$\forall i \in C \setminus C_R \cup C_S, \quad t \in \mathfrak{T},$$

$$\sum_{j \in C} a_{ij} y_{ij}^t \leq Q_i^t + \phi_i b_i$$

$$\forall i \in C \setminus C_S, \quad t \in \mathfrak{T},$$

$$x_i^0 = \hat{x}_i \quad \forall i \in C \setminus C_S,$$

$$y_{ij}^0 = 0, \quad \forall (i, j) \in C \times C,$$

$$x_i^t \geq 0, \quad \forall i \in C \setminus C_S, \quad t \in \mathfrak{T},$$

$$\begin{aligned}
y_{ij}^t &\geq 0, \quad \forall (i, j) \in C \times C, t \in \mathfrak{F}, \\
b_i &\geq 0, \quad \forall i \in C \setminus C_S, \\
\sum_{i \in C \setminus C_S} b_i &\leq B.
\end{aligned} \tag{19}$$

Ben-Tal et al. [22] showed that the ARC is more flexible than the RC; however, in most cases the ARC is computationally intractable (NP-hard). To address this difficulty, Ben-Tal et al. [22] proposed the affine decision rules; that is, adjustable variables are restricted to be affine function of the uncertain data. Here, we use the affine decision rules proposed by Ben-Tal et al. [22] to make ARC computationally tractable. Then,

$$\begin{aligned}
x_i^t &= \eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^\tau, \\
y_{ij}^t &= \pi_{ijt}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \pi_{ijt}^{s\tau} d_s^\tau,
\end{aligned} \tag{20}$$

where  $y_{ij}^t$  is adjustable control variable,  $x_i^t$  is state variable,  $\eta_{it}^{-1}$ ,  $\pi_{ijt}^{-1}$ ,  $\eta_{it}^{s\tau}$ , and  $\pi_{ijt}^{s\tau}$  are nonadjustable variables, and  $I_t = \{0, 1, 2, \dots, t-1\}$ . By substituting the state and control variable, we have the following SODNP-AARC formulation:

SODNP-AARC:

$$\begin{aligned}
&\min_{\eta, \pi, b, z} z, \\
&\text{s.t.} \quad \sum_{t \in \mathfrak{F}} \sum_{i \in C \setminus C_S} c_i^t \left( \eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^\tau \right) \leq z, \quad \forall d_s^\tau \in U, \\
&\quad \left( \eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^\tau \right) \\
&\quad - \left( \eta_{it-1}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_{t-1}} \eta_{it-1}^{s\tau} d_s^\tau \right) \\
&\quad - \sum_{k \in C} a_{ki} \left( \pi_{kit-1}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_{t-1}} \pi_{kit-1}^{s\tau} d_s^\tau \right) \\
&\quad + \sum_{j \in C} a_{ij} \left( \pi_{ijt}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \pi_{ijt}^{s\tau} d_s^\tau \right) \geq d_i^{t-1}, \\
&\quad \forall i \in C_R, t \in \mathfrak{F}, d_i^t \in U, \\
&\quad \left( \eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^\tau \right) \\
&\quad - \left( \eta_{it-1}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_{t-1}} \eta_{it-1}^{s\tau} d_s^\tau \right) \\
&\quad - \sum_{k \in C} a_{ki} \left( \pi_{kit-1}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_{t-1}} \pi_{kit-1}^{s\tau} d_s^\tau \right)
\end{aligned}$$

$$+ \sum_{j \in C} a_{ij} \left( \pi_{ijt}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \pi_{ijt}^{s\tau} d_s^\tau \right) \geq 0,$$

$$\forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, d_i^t \in U,$$

$$\sum_{j \in C} a_{ij} \left( \pi_{ijt}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \pi_{ijt}^{s\tau} d_s^\tau \right)$$

$$\leq \left( \eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^\tau \right),$$

$$\forall i \in C \setminus C_S, t \in \mathfrak{F}, d_i^t \in U,$$

$$\sum_{k \in C} a_{ki} \left( \pi_{kit}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \pi_{kit}^{s\tau} d_s^\tau \right)$$

$$\leq \delta_i^t \left( N_i^t + \chi_i b_i - \left( \eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^\tau \right) \right),$$

$$\forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, d_i^t \in U,$$

$$\sum_{k \in C} a_{ki} \left( \pi_{kit}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \pi_{kit}^{s\tau} d_s^\tau \right) \leq Q_i^t + \phi_i b_i,$$

$$\forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, d_i^t \in U,$$

$$\sum_{k \in C} a_{ki} \left( \pi_{kit}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \pi_{kit}^{s\tau} d_s^\tau \right)$$

$$\leq Q_i^t + \phi_i b_i, \quad \forall i \in C \setminus C_S, t \in \mathfrak{F}, d_i^t \in U,$$

$$\eta_{i0}^{-1} = \hat{x}_i, \quad \forall i \in C \setminus C_S,$$

$$\pi_{ij0}^{-1} = 0, \quad \forall (i, j) \in C \times C,$$

$$\eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^\tau \geq 0,$$

$$\forall i \in C \setminus C_S, t \in \mathfrak{F}, \bar{d}_i^t \in U,$$

$$\pi_{ijt}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \pi_{ijt}^{s\tau} d_s^\tau \geq 0,$$

$$\forall (i, j) \in C \times C, t \in \mathfrak{F}, \bar{d}_i^t \in U,$$

$$\sum_{i \in C \setminus C_S} b_i \leq B,$$

$$b_i \geq 0, \quad \forall i \in C \setminus C_S.$$

(21)

It can be seen from the above formulation that SODNP-AARC is intractable since it is a semi-infinite program, and it can be reformulated as a tractable optimization problem as shown in Theorem 1.

**Theorem 1.** Given a polyhedral uncertainty set  $U$ , the affinely adjustable robust counterpart of the SODNDP becomes the following linear programming and thus computationally tractable. Note that  $\lambda$  is a set of dual variables and the numerical indexes are used for notational simplicity:

SODNDP-AARCI:

$$\min_{b,z,\lambda,\pi,\eta} z, \quad (22)$$

$$\text{s.t.} \quad \sum_{t \in \mathfrak{S}} \sum_{i \in C \setminus C_s} c_i^t \eta_{it}^{-1} - \sum_{s \in C_R} \sum_{\tau \in \mathfrak{S}} (1 - \theta) \lambda_{st}^{11} \bar{d}_s^\tau \quad (23)$$

$$+ \sum_{s \in C_R} \sum_{\tau \in \mathfrak{S}} (1 + \theta) \lambda_{st}^{12} \bar{d}_s^\tau + \sum_{s \in C_R} \lambda_s^{13} D_s \leq z,$$

$$\lambda_{st}^{12} - \lambda_{st}^{11} + \lambda_s^{13} = \sum_{i \in C \setminus C_s} \sum_{t=\{\tau+1, \dots, T\}} c_i^t \eta_{it}^{st}, \quad (24)$$

$$\forall s \in C_R, \tau \in \{0, 1, 2, \dots, T-1\},$$

$$\lambda_{st}^{12} - \lambda_{st}^{11} + \lambda_s^{13} = 0, \quad \forall s \in C_R, \tau \in \{T\}, \quad (25)$$

$$\lambda_{st}^{12}, \lambda_{st}^{11}, \lambda_s^{13} \geq 0, \quad \forall s \in C_R, \tau \in \mathfrak{S}, \quad (26)$$

$$\eta_{it}^{-1} - \eta_{it-1}^{-1} - \sum_{k \in C} a_{ki} \pi_{kit-1}^{-1} + \sum_{j \in C} a_{ij} \pi_{ijt-1}^{-1} \quad (27)$$

$$+ \sum_{s \in C_R, \tau \in \mathfrak{S}} \lambda_{its\tau}^{21} (1 - \theta) \bar{d}_s^\tau$$

$$- \sum_{s \in C_R, \tau \in \mathfrak{S}} \lambda_{its\tau}^{22} (1 + \theta) \bar{d}_s^\tau$$

$$- \sum_{s \in C_R} \lambda_{its}^{23} D_s \geq 0, \quad \forall i \in C_R, t \in \mathfrak{S},$$

$$\lambda_{its\tau}^{22} - \lambda_{its\tau}^{21} + \lambda_{its}^{23} = I_{\{\tau=t-1, s=i\}} - \eta_{it}^{st} \quad (28)$$

$$+ \left( \eta_{it-1}^{st} + \sum_{k \in C} a_{ki} \pi_{kit-1}^{st} - \sum_{j \in C} a_{ij} \pi_{ijt-1}^{st} \right)$$

$$\times I_{\{\tau < t-1\}}, \quad \forall i \in C_R, t \in \mathfrak{S},$$

$$s \in C_R, \tau \in \{0, 1, \dots, t-1\}, \quad (29)$$

$$\lambda_{its\tau}^{22} - \lambda_{its\tau}^{21} + \lambda_{its}^{23} = 0,$$

$$\forall i \in C_R, t \in \mathfrak{S}, s \in C_R, \tau \in \{t, \dots, T\},$$

$$\lambda_{its\tau}^{21}, \lambda_{its\tau}^{22}, \lambda_{its}^{23} \geq 0, \quad (30)$$

$$\forall i \in C_R, t \in \mathfrak{S}, s \in C_R, \tau \in \mathfrak{S},$$

$$\eta_{it}^{-1} - \eta_{it-1}^{-1} - \sum_{k \in C} a_{ki} \pi_{kit-1}^{-1} + \sum_{j \in C} a_{ij} \pi_{ijt-1}^{-1} \quad (31)$$

$$+ \sum_{s \in C_R, \tau \in \mathfrak{S}} \lambda_{its\tau}^{31} (1 - \theta) \bar{d}_s^\tau - \sum_{s \in C_R, \tau \in \mathfrak{S}} \lambda_{its\tau}^{32} (1 + \theta) \bar{d}_s^\tau$$

$$- \sum_{s \in C_R} \lambda_{its}^{33} D_s \geq 0, \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{S},$$

$$\lambda_{its\tau}^{32} - \lambda_{its\tau}^{31} + \lambda_{its}^{33} = -\eta_{it}^{st} \quad (32)$$

$$+ \left( \eta_{it-1}^{st} + \sum_{k \in C} a_{ki} \pi_{kit-1}^{st} - \sum_{j \in C} a_{ij} \pi_{ijt-1}^{st} \right)$$

$$\times I_{\{\tau < t-1\}}, \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{S},$$

$$s \in C_R, \tau \in \{0, 1, \dots, t-1\},$$

$$\lambda_{its\tau}^{32} - \lambda_{its\tau}^{31} + \lambda_{its}^{33} = 0, \quad \forall i \in C \setminus C_R \cup C_S, \quad (33)$$

$$t \in \mathfrak{S}, s \in C_R, \tau \in \{t, \dots, T\},$$

$$\lambda_{its\tau}^{31}, \lambda_{its\tau}^{32}, \lambda_{its}^{33} \geq 0, \quad \forall i \in C \setminus C_R \cup C_S, \quad (34)$$

$$t \in \mathfrak{S}, s \in C_R, \tau \in \mathfrak{S},$$

$$\sum_{j \in C} a_{ij} \pi_{ijt}^{-1} - \eta_{it}^{-1} + \sum_{s \in C_R, \tau \in \mathfrak{S}} \lambda_{its\tau}^{42} (1 + \theta) \bar{d}_s^\tau \quad (35)$$

$$- \sum_{s \in C_R, \tau \in \mathfrak{S}} \lambda_{its\tau}^{41} (1 - \theta) \bar{d}_s^\tau + \sum_{s \in C_R} \lambda_{its}^{43} D_s \leq 0,$$

$$\forall i \in C \setminus C_S, t \in \mathfrak{S},$$

$$\lambda_{its\tau}^{42} - \lambda_{its\tau}^{41} + \lambda_{its}^{43} = \sum_{j \in C} a_{ij} \pi_{ijt}^{st} - \eta_{it}^{st}, \quad (36)$$

$$\forall i \in C \setminus C_S, t \in \mathfrak{S}, s \in C_R, \tau \in \{0, 1, \dots, t-1\},$$

$$\lambda_{its\tau}^{42} - \lambda_{its\tau}^{41} + \lambda_{its}^{43} = 0, \quad \forall i \in C \setminus C_S, \quad (37)$$

$$t \in \mathfrak{S}, s \in C_R, \tau \in \{t, \dots, T\},$$

$$\lambda_{its\tau}^{41}, \lambda_{its\tau}^{42}, \lambda_{its}^{43} \geq 0, \quad \forall i \in C \setminus C_S, \quad (38)$$

$$t \in \mathfrak{S}, s \in C_R, \tau \in \mathfrak{S},$$

$$\sum_{k \in C} a_{ki} \pi_{kit}^{-1} + \sum_{s \in C_R, \tau \in \mathfrak{S}} \lambda_{its\tau}^{52} (1 + \theta) \bar{d}_s^\tau \quad (39)$$

$$- \sum_{s \in C_R, \tau \in \mathfrak{S}} \lambda_{its\tau}^{51} (1 - \theta) \bar{d}_s^\tau$$

$$+ \sum_{s \in C_R} \lambda_{its}^{53} D_s \leq \delta_i^t (N_i^t + \chi_i b_i - \eta_{it}^{-1}),$$

$$\forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{S},$$

$$\lambda_{its\tau}^{52} - \lambda_{its\tau}^{51} + \lambda_{its}^{53} = \sum_{k \in C} a_{ki} \pi_{kit}^{st} + \delta_i^t \eta_{it}^{st}, \quad (40)$$

$$\forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{S},$$

$$s \in C_R, \tau \in \{0, 1, \dots, t-1\},$$

$$\lambda_{its\tau}^{52} - \lambda_{its\tau}^{51} + \lambda_{its}^{53} = 0, \quad \forall i \in C \setminus C_R \cup C_S, \quad (41)$$

$$t \in \mathfrak{S}, s \in C_R, \tau \in \{t, \dots, T\},$$

$$\lambda_{its\tau}^{51}, \lambda_{its\tau}^{52}, \lambda_{its}^{53} \geq 0, \quad \forall i \in C \setminus C_R \cup C_S, \quad (42)$$

$$t \in \mathfrak{S}, s \in C_R, \tau \in \mathfrak{S},$$

$$\begin{aligned} & \sum_{k \in C} a_{ki} \pi_{kit}^{-1} + \sum_{s \in C_R, \tau \in \mathfrak{F}} \lambda_{its\tau}^{62} (1 + \theta) \bar{d}_s^T \\ & - \sum_{s \in C_R, \tau \in \mathfrak{F}} \lambda_{its\tau}^{61} (1 - \theta) \bar{d}_s^T \end{aligned} \quad (43)$$

$$\begin{aligned} & + \sum_{s \in C_R} \lambda_{its}^{63} D_s \leq Q_i + \phi_i b_i, \\ & \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, \\ & \lambda_{its\tau}^{62} - \lambda_{its\tau}^{61} + \lambda_{its}^{63} = \sum_{k \in C} a_{ki} \pi_{kit}^{s\tau}, \\ & \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, \end{aligned} \quad (44)$$

$$\begin{aligned} & s \in C_R, \tau \in \{0, 1, \dots, t-1\}, \\ & \lambda_{its\tau}^{62} - \lambda_{its\tau}^{61} + \lambda_{its}^{63} = 0, \quad \forall i \in C \setminus C_R \cup C_S, \\ & t \in \mathfrak{F}, s \in C_R, \tau \in \{t, \dots, T\}, \end{aligned} \quad (45)$$

$$\begin{aligned} & \lambda_{its\tau}^{61}, \lambda_{its\tau}^{62}, \lambda_{its}^{63} \geq 0, \quad \forall i \in C \setminus C_R \cup C_S, \\ & t \in \mathfrak{F}, s \in C_R, \tau \in \mathfrak{F}, \end{aligned} \quad (46)$$

$$\begin{aligned} & \sum_{j \in C} a_{ij} \pi_{ijt}^{-1} + \sum_{s \in C_R, \tau \in \mathfrak{F}} \lambda_{its\tau}^{72} (1 + \theta) \bar{d}_s^T \\ & - \sum_{s \in C_R, \tau \in \mathfrak{F}} \lambda_{its\tau}^{71} (1 - \theta) \bar{d}_s^T \end{aligned} \quad (47)$$

$$\begin{aligned} & + \sum_{s \in C_R} \lambda_{its}^{73} D_s \leq Q_i + \phi_i b_i, \\ & \forall i \in C \setminus C_S, t \in \mathfrak{F}, \\ & \lambda_{its\tau}^{72} - \lambda_{its\tau}^{71} + \lambda_{its}^{73} = \sum_{j \in C} a_{ij} \pi_{ijt}^{s\tau}, \\ & \forall i \in C \setminus C_S, t \in \mathfrak{F}, \end{aligned} \quad (48)$$

$$\begin{aligned} & s \in C_R, \tau \in \{0, 1, \dots, t-1\}, \\ & \lambda_{its\tau}^{72} - \lambda_{its\tau}^{71} + \lambda_{its}^{73} = 0, \quad \forall i \in C \setminus C_S, t \in \mathfrak{F}, \\ & s \in C_R, \tau \in \{t, \dots, T\}, \end{aligned} \quad (49)$$

$$\begin{aligned} & \lambda_{its\tau}^{71}, \lambda_{its\tau}^{72}, \lambda_{its}^{73} \geq 0, \\ & \forall i \in C \setminus C_S, t \in \mathfrak{F}, s \in C_R, \tau \in \mathfrak{F}, \end{aligned} \quad (50)$$

$$\begin{aligned} & \eta_{it}^{-1} - \sum_{s \in C_R, \tau \in \mathfrak{F}} \lambda_{its\tau}^{82} (1 + \theta) \bar{d}_s^T \\ & + \sum_{s \in C_R, \tau \in \mathfrak{F}} \lambda_{its\tau}^{81} (1 - \theta) \bar{d}_s^T \end{aligned} \quad (51)$$

$$\begin{aligned} & - \sum_{s \in C_R} \lambda_{its}^{83} D_s \geq 0, \quad \forall i \in C \setminus C_S, t \in \mathfrak{F}, \\ & \lambda_{its\tau}^{82} - \lambda_{its\tau}^{81} + \lambda_{its}^{83} = -\eta_{it}^{s\tau}, \quad \forall i \in C \setminus C_S, \\ & t \in \mathfrak{F}, s \in C_R, \tau \in \{0, 1, \dots, t-1\}, \end{aligned} \quad (52)$$

$$\begin{aligned} & \lambda_{its\tau}^{82} - \lambda_{its\tau}^{81} + \lambda_{its}^{83} = 0, \quad \forall i \in C \setminus C_S, \\ & t \in \mathfrak{F}, s \in C_R, \tau \in \{t, \dots, T\}, \end{aligned} \quad (53)$$

$$\begin{aligned} & \lambda_{its\tau}^{81}, \lambda_{its\tau}^{82}, \lambda_{its}^{83} \geq 0, \\ & \forall i \in C \setminus C_S, t \in \mathfrak{F}, s \in C_R, \tau \in \mathfrak{F}, \end{aligned} \quad (54)$$

$$\begin{aligned} & \pi_{ijt}^{-1} - \sum_{s \in C_R, \tau \in \mathfrak{F}} \lambda_{ijts\tau}^{92} (1 + \theta) \bar{d}_s^T \\ & + \sum_{s \in C_R, \tau \in \mathfrak{F}} \lambda_{ijts\tau}^{91} (1 - \theta) \bar{d}_s^T - \sum_{s \in C_R} \lambda_{ijts}^{93} D_s \geq 0, \\ & \forall (i, j) \in \{a(i, j) = 1\}, t \in \mathfrak{F}, \end{aligned} \quad (55)$$

$$\begin{aligned} & \lambda_{ijts\tau}^{92} - \lambda_{ijts\tau}^{91} + \lambda_{ijts}^{93} = -\pi_{ijt}^{s\tau}, \quad \forall (i, j) \in \{a(i, j) = 1\}, \\ & t \in \mathfrak{F}, s \in C_R, \tau \in \{0, 1, \dots, t-1\}, \end{aligned} \quad (56)$$

$$\begin{aligned} & \lambda_{ijts\tau}^{92} - \lambda_{ijts\tau}^{91} + \lambda_{ijts}^{93} = 0, \quad \forall (i, j) \in \{a(i, j) = 1\}, \\ & t \in \mathfrak{F}, s \in C_R, \tau \in \{t, \dots, T\}, \end{aligned} \quad (57)$$

$$\begin{aligned} & \lambda_{ijts\tau}^{91}, \lambda_{ijts\tau}^{92}, \lambda_{ijts}^{93} \geq 0, \quad \forall (i, j) \in \{a(i, j) = 1\}, \\ & t \in \mathfrak{F}, s \in C_R, \tau \in \mathfrak{F}, \end{aligned} \quad (58)$$

$$\eta_{i0}^{-1} = \hat{x}_i, \quad \forall i \in C \setminus C_S, \quad (59)$$

$$\pi_{ij0}^{-1} = 0, \quad \forall (i, j) \in \{a(i, j) = 1\}, \quad (60)$$

$$\begin{aligned} & \sum_{i \in C \setminus C_S} b_i \leq B, \\ & b_i \geq 0, \quad \forall i \in C \setminus C_S, \end{aligned} \quad (61)$$

where  $I_{\{\tau=t-1, s=i\}}$  in the constraint (28) denotes that if  $\tau = t - 1$ ,  $s = i$ , then  $I_{\{\tau=t-1, s=i\}} = 1$ ; otherwise  $I_{\{\tau=t-1, s=i\}} = 0$ , and  $I_{\{\tau < t-1\}}$  in the constraint (28) denotes that if  $\tau < t - 1$ , then  $I_{\{\tau < t-1\}} = 1$ ; otherwise  $I_{\{\tau < t-1\}} = 0$ .

*Proof.* By applying the following relationship, we can reformulate each constraint affected by demand uncertainty as an equivalent LP problem. For example,

$$\sum_{t \in \mathfrak{F}} \sum_{i \in C \setminus C_S} c_i^t \left( \eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^T \right) \leq z,$$

$$\forall d_s^T \in U = \left\{ (1 - \theta) \bar{d}_s^T \leq d_s^T \leq (1 + \theta) \bar{d}_s^T, \forall s \in C_R, \tau \in \mathfrak{F}, \right.$$

$$\left. \sum_{\tau \in \mathfrak{F}} d_s^T \leq D_s, \forall s \in C_R \right\}$$

$$\iff \max_{d_s^T \in U} \sum_{t \in \mathfrak{F}} \sum_{i \in C \setminus C_S} c_i^t \left( \eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^T \right) \leq z. \quad (62)$$

According to strong duality property, we can derive the equivalent constraint with dual problem of the above problem:

$$\begin{aligned}
& \max_{d_s^t} \sum_{t \in \mathfrak{F}} \sum_{i \in C \setminus C_s} c_i^t \left( \eta_{it}^{-1} + \sum_{s \in C_R} \sum_{\tau \in I_t} \eta_{it}^{s\tau} d_s^\tau \right) \leq z \\
& \text{s.t.} \quad (1 - \theta) \bar{d}_s^\tau \leq d_s^\tau \leq (1 + \theta) \bar{d}_s^\tau, \quad \forall s \in C_R, \tau \in \mathfrak{F}, \\
& \quad \sum_{\tau \in \mathfrak{F}} d_s^\tau \leq D_s, \quad \forall s \in C_R \\
& \quad \Downarrow \\
& \min_{\eta_{it}^{-1}, \lambda_{st}^{11}, \lambda_{st}^{12}, \lambda_s^{13}} \sum_{i \in C \setminus C_s, t \in \mathfrak{F}} c_i^t \eta_{it}^{-1} - \sum_{s \in C_R, \tau \in \mathfrak{F}} (1 - \theta) \lambda_{st}^{11} \bar{d}_s^\tau \\
& \quad + \sum_{s \in C_R, \tau \in \mathfrak{F}} (1 + \theta) \lambda_{st}^{12} \bar{d}_s^\tau + \sum_{s \in C_R} \lambda_s^{13} D_s \leq z, \\
& \text{s.t.} \quad \lambda_{st}^{12} - \lambda_{st}^{11} + \lambda_s^{13} = \sum_{i \in C \setminus C_s} \sum_{t = \{\tau+1, \dots, T\}} c_i^t \eta_{it}^{s\tau}, \\
& \quad \forall s \in C_R, \tau \in \{0, 1, 2, \dots, T-1\}, \\
& \quad \lambda_{st}^{12} - \lambda_{st}^{11} + \lambda_s^{13} = 0, \quad \forall s \in C_R, \tau \in \{T\}, \\
& \quad \lambda_{st}^{12}, \lambda_{st}^{11}, \lambda_s^{13} \geq 0, \quad \forall s \in C_R, \tau \in \mathfrak{F},
\end{aligned} \tag{63}$$

where  $\lambda_{st}^{11}$ ,  $\lambda_{st}^{12}$ , and  $\lambda_s^{13}$  are dual variables. Therefore, we can derive the constraints (23)–(26). Similarity, the other semi-infinite constraints can also write as the equivalent linear constraints.  $\square$

Note that the optimal objective of SODNDP-AARCI is a guaranteed upper bound value for all realization of uncertain data under the assumption of linear dependency. It also can be interpreted as the optimistic estimate of total cost in worst case, which can be lower than the optimistic estimate from robust counterpart as SODNDP-AARCI has a larger robust feasible region [22]. Moreover, the decision variables of SODNDP-AARCI are not adjustable control and state variables but a set of coefficient of affine function of the control variables including  $\pi_{ijt}^{-1}$ ,  $\pi_{ijt}^{s\tau}$ ,  $\eta_{it}^{-1}$ , and  $\eta_{it}^{s\tau}$ . It implies that the solution of SODNDP-AARCI is a linear decision rule. Specific values of  $x_i^t$  and  $y_{ij}^t$  are computed after the realization of the demand at time  $t - 1$  [34].

## 5. Robust Formulation

To demonstrate that the AARC is more flexible, we use the usual RO approach to develop the RC of SODNDP in this section. We note that constraint (11) is redundant since constraints (5) and (12) guarantee the nonnegativity of  $x_i^t$ . By

using constraints (3), (4), (9), and (10),  $x_i^t$  can be represented as follows:

$$\begin{aligned}
x_i^t &= \hat{x}_i + \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau + d_i^\tau \right), \\
& \quad \forall i \in C_R, t \in \mathfrak{S}, \\
x_i^t &= \hat{x}_i + \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau \right), \\
& \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}.
\end{aligned} \tag{64}$$

As before, it is also assumed that the demand uncertainty belongs to the uncertain set  $U$ . Thus, we can get the following RC of SODNDP:

SODNDP-RC:

$$\begin{aligned}
& \min_{b, y, z} z, \\
& \text{s.t.} \quad \sum_{t \in \mathfrak{F}} \sum_{i \in C_R} c_i^t \\
& \quad \times \left( \hat{x}_i + \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau + d_i^\tau \right) \right) \\
& \quad + \sum_{t \in \mathfrak{F}} \sum_{i \in C \setminus C_R \cup C_S} c_i^t \\
& \quad \times \left( \hat{x}_i + \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau \right) \right) \\
& \leq z, \quad \forall d_i^t \in U, \\
& \quad \sum_{j \in C} a_{ij} y_{ij}^t \\
& \leq \hat{x}_i + \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau + d_i^\tau \right), \\
& \quad \forall i \in C_R, t \in \mathfrak{F}, d_i^t \in U, \\
& \quad \sum_{j \in C} a_{ij} y_{ij}^t \\
& \leq \hat{x}_i + \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau \right), \\
& \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F},
\end{aligned}$$

$$\begin{aligned}
& \sum_{k \in C} a_{ki} y_{ki}^t \\
& \leq \delta_i^t \left( N_i^t + \chi_i b_i - \hat{x}_i \right. \\
& \quad \left. - \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau \right) \right), \\
& \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, \\
& \sum_{k \in C} a_{ki} y_{ki}^t \leq Q_i^t + \phi_i b_i, \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, \\
& \sum_{j \in C} a_{ij} y_{ij}^t \leq Q_i^t + \phi_i b_i, \quad \forall i \in C \setminus C_S, t \in \mathfrak{F}, \\
& y_{ij}^0 = 0, \quad \forall (i, j) \in C \times C, \\
& y_{ij}^t \geq 0, \quad \forall (i, j) \in C \times C, t \in \mathfrak{F}, \\
& \sum_{i \in C \setminus C_S} b_i \leq B, \\
& b_i \geq 0, \quad \forall i \in C \setminus C_S,
\end{aligned} \tag{65}$$

It is obvious that RC is also a semi-infinite problem and has infinitely many constraints. Similar to Theorem 1, we can convert it into the following tractable equivalent deterministic problem:

SODNDP-RC1:

$$\begin{aligned}
& \min_{b, y, \mu, z} z, \\
& \text{s.t.} \quad \sum_{t \in \mathfrak{F}} \sum_{i \in C \setminus C_S} c_i^t \\
& \quad \times \left( \hat{x}_i + \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau \right) \right) \\
& \quad - \sum_{s \in C_R, \tau \in \mathfrak{F}} \mu_{s\tau}^{11} (1 - \theta) \bar{d}_s^\tau + \sum_{s \in C_R, \tau \in \mathfrak{F}} \mu_{s\tau}^{12} (1 + \theta) \bar{d}_s^\tau \\
& \quad + \sum_{s \in C_R} \mu_s^{13} D_s \leq z, \\
& \mu_{s\tau}^{12} - \mu_{s\tau}^{11} + \mu_s^{13} = \sum_{t=\{\tau+1, \dots, T\}} c_s^t, \\
& \quad \forall s \in C_R, \tau \in \{0, 1, 2, \dots, T-1\}, \\
& \mu_{s\tau}^{12} - \mu_{s\tau}^{11} + \mu_s^{13} = 0, \quad \forall s \in C_R, \tau \in \{T\}, \\
& \mu_{s\tau}^{11}, \mu_{s\tau}^{12}, \mu_s^{13} \geq 0, \quad \forall s \in C_R, \tau \in \mathfrak{F},
\end{aligned}$$

$$\begin{aligned}
& \sum_{j \in C} a_{ij} y_{ij}^t \\
& \leq \hat{x}_i + \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau \right) \\
& \quad + \sum_{s \in C_R} \sum_{\tau \in \mathfrak{F}} \mu_{its\tau}^{21} (1 - \theta) \bar{d}_s^\tau \\
& \quad - \sum_{s \in C_R} \sum_{\tau \in \mathfrak{F}} \mu_{its\tau}^{22} (1 + \theta) \bar{d}_s^\tau \\
& \quad - \sum_{s \in C_R} \mu_{its}^{23} D_s, \quad \forall i \in C_R, t \in \mathfrak{F}, \\
& \mu_{its\tau}^{22} - \mu_{its\tau}^{21} + \mu_{its}^{23} + I_{\{i=s, \tau < t\}} = 0, \\
& \quad \forall i \in C_R, t \in \mathfrak{F}, s \in C_R, \tau \in \mathfrak{F}, \\
& \mu_{its\tau}^{21}, \mu_{its\tau}^{22}, \mu_{its}^{23} \geq 0, \quad \forall i \in C_R, t \in \mathfrak{F}, s \in C_R, \tau \in \mathfrak{F}, \\
& \sum_{j \in C} a_{ij} y_{ij}^t \\
& \leq \hat{x}_i + \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau \right), \\
& \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, \\
& \sum_{k \in C} a_{ki} y_{ki}^t \\
& \leq \delta_i^t \left( N_i^t + \chi_i b_i - \hat{x}_i \right. \\
& \quad \left. - \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau \right) \right) \\
& \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, \\
& \sum_{k \in C} a_{ki} y_{ki}^t \\
& \leq \delta_i^t \left( N_i^t + \chi_i b_i - \hat{x}_i \right. \\
& \quad \left. - \sum_{\tau=0}^{t-1} \left( \sum_{k \in C} a_{ki} y_{ki}^\tau - \sum_{j \in C} a_{ij} y_{ij}^\tau \right) \right) \\
& \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, \\
& \sum_{k \in C} a_{ki} y_{ki}^t \leq Q_i^t + \phi_i b_i, \quad \forall i \in C \setminus C_R \cup C_S, t \in \mathfrak{F}, \\
& \sum_{j \in C} a_{ij} y_{ij}^t \leq Q_i^t + \phi_i b_i, \quad \forall i \in C \setminus C_S, t \in \mathfrak{F}, \\
& y_{ij}^0 = 0, \quad \forall (i, j) \in C \times C, \\
& y_{ij}^t \geq 0, \quad \forall (i, j) \in C \times C, t \in \mathfrak{F}, \\
& \sum_{i \in C \setminus C_S} b_i \leq B, \\
& b_i \geq 0, \quad \forall i \in C \setminus C_S,
\end{aligned} \tag{66}$$

where  $\mu$  is a set of dual variables and the numerical indexes are used for notational simplicity.

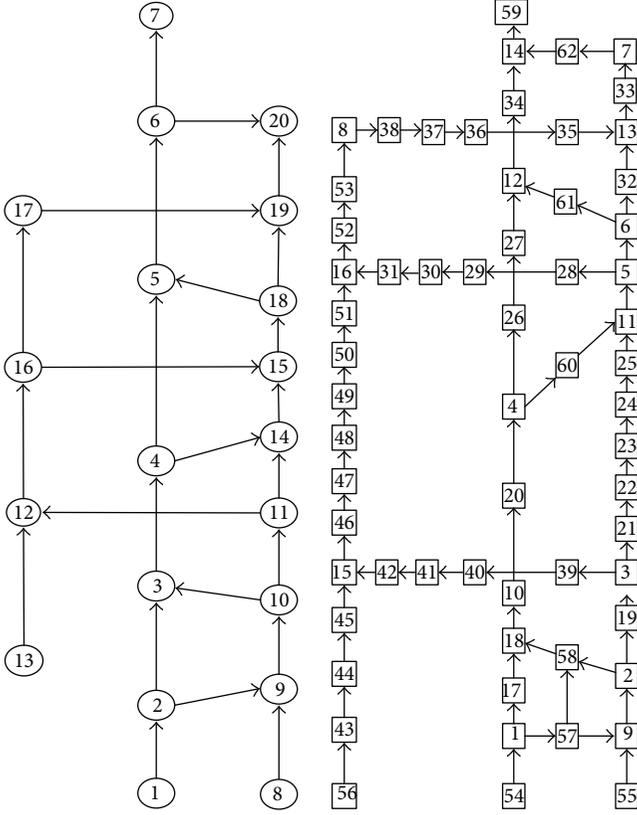


FIGURE 1: Test network [40].

TABLE 1: Characteristic of the cells.

Cell	Source cells	Sink cells	Freeway cells	Arterial cells
$N_i^t$	$\infty$	$\infty$	20	10
$Q_i^t$	$\infty$	$\infty$	12	3
$\delta_i^t$	1	1	1	1
$\tilde{x}_i$	0	0	0	0

## 6. Numerical Example

The purpose of presenting the numerical example in this section is twofold: (1) to demonstrate the validity of SODNDP-AARC1 and (2) to illustrate the flexibility of the SODNDP-AARC1 by comparing it with SODNDP-RC1. One example network shown in Figure 1 is selected to test the proposed model. This network consists of 20 nodes and 25 links. The links in the center of the network are the freeway links, while the outer link and crosslink are the arterial links. The characteristics of network include length, speed limit, and jam density of the roads. By using this network information and setting 6 seconds as the length of a time interval, the corresponding cell network representation is constructed. This cell network is composed of 62 cells and 67 cell connectors. There are three source cells (cells 54, 55, and 56) and one sink cell (cell 59). The other is either freeway cell or arterial cells. Expected the sink cell 59, the other cells are considered for capacity expansion. The characteristics of the cells in the test network are shown in Table 1.

The planning horizon  $T$  and cost coefficient  $c_i^t$  except the one at the end of planning horizon are assumed to be 105 and 1, respectively.  $c_i^T$  is set to 10 to penalize the unmet demand. The model parameters  $\delta$  and  $\phi$  are assumed to be unity; that is,  $\delta_i^t = \phi_i = 1$ , for all  $i \in C, t \in \mathfrak{T}$ . Travel demand is only generated at the three source cells and time 0. The nominal demand is set to 63.6. That is,  $d_i^0 = 63.6$ , for all  $i \in C_R$ ; the total demand is  $D_i = 2 \sum d_i^t$ , for all  $i \in C_R$ . The SODNDP-AARC1 and SODNDP-RC1 problems are solved with GUROBI using GAMS [37] on a PC with Intel processor 2.8 GHz and 4 GB of memory.

Table 2 reports the optimal solution of the SODNDP, SODNDP-AARC1, and SODNDP-RC1 models under different uncertainty levels when the available budget  $B = 100$ . It can be seen from Table 2 that the optimal solution of the SODNDP-AARC1 model has remained unchanged under different uncertainty levels, while the optimal solution of the SODNDP-RC1 model keeps changing. It is clear that the adjustable robust capacity expansion plans are more stable. Tables 3 and 4 list the total travel cost and solution time of the SODNDP, SODNDP-AARC1, and SODNDP-RC1 models under different uncertainty levels and available budgets. As shown in Tables 3 and 4, the total travel cost from the nominal solution, robust solution, and adjustable robust solution increases with the rise of uncertainty level  $\theta$  when  $B = 100$ . On the other hand, the total travel cost from the nominal solution, robust solution, and adjustable robust solution decreases with the increase of available budget  $B$  when  $\theta = 0.5$ . This is because the feasible region of all three models reduces with the rise of uncertainty level  $\theta$ ; on the contrary, the feasible region of all three models enlarges with the increase of available budget  $B$ . At the same time, the optimal objective value of the SODNDP-RC1 model is far greater not only than the SODNDP model but also than the SODNDP-AARC1 model, while the optimal objective value of SODNDP-AARC1 model is only slightly greater than SODNDP model. This shows that the SODNDP-RC1 model is overly conservative, while SODNDP-AARC1 model is more flexible than the SODNDP-RC1 model. Moreover, the solution time of the SODNDP model and SODNDP-RC1 model is almost the same, and the solution time of the SODNDP-AARC1 model is longer than the other two models. It is due to the fact that the SODNDP-AARC1 model has more variables and constraints.

Figure 2 depicts the total travel cost from the nominal solution, robust solution, and adjustable solution and the relative improvement of the total travel cost of the robust solution against adjustable solution under different uncertainty levels and available budgets. The relative improvement in this paper is defined as

$$RI = \frac{z^R - z^{AR}}{z^R}, \quad (67)$$

where  $z^R$  is the total travel cost from the robust solution and  $z^{AR}$  is the total travel cost from the adjustable robust solution. As shown in Figure 2, the total travel cost from both the robust solution and adjustable robust solution decreases with the increase of available budget  $B$  but increases with

TABLE 2: The optimal solution of the SODNDP, SODNDP-RCI, and SODNDP-AARCI when  $B = 100$ .

Cell	SODNDP										
$\theta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	0.72	0	0
2	0.762	2.082	3.402	5.235	7.492	5.9	9.539	9.08	9.72	3.36	0
4	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	5.44	0	0
5	0	0	0	0	0	0	0	0	0	0	100
9	0.762	2.082	3.402	5.235	7.492	5.9	9.539	9.08	9.72	3.36	0
10	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	5.44	0	0
12	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	5.44	0	0
14	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	5.44	0	0
17	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	0.72	0	0
18	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	5.44	0	0
20	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	5.44	0	0
26	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	25.88	0	0
27	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	5.44	0	0
34	8.883	8.523	8.163	7.663	7.048	7.482	6.489	6.615	5.44	89.92	0
58	0.762	2.082	3.402	5.235	7.492	5.9	9.539	9.08	9.72	3.36	0

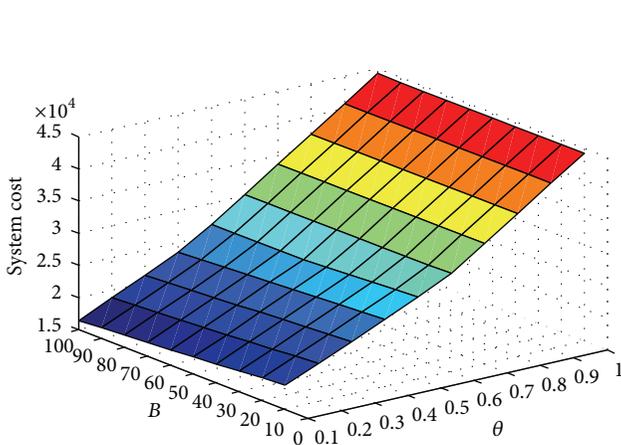
Cell	SODNDP-AARCI										
$\theta$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
1	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
4	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
10	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
12	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
14	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
17	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
18	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
20	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
26	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
27	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091
34	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091	9.091

TABLE 3: The optimal objective and solution time of the SODNDP, SODNDP-RCI, and SODNDP-AARCI models under different uncertainty levels when  $B = 100$ .

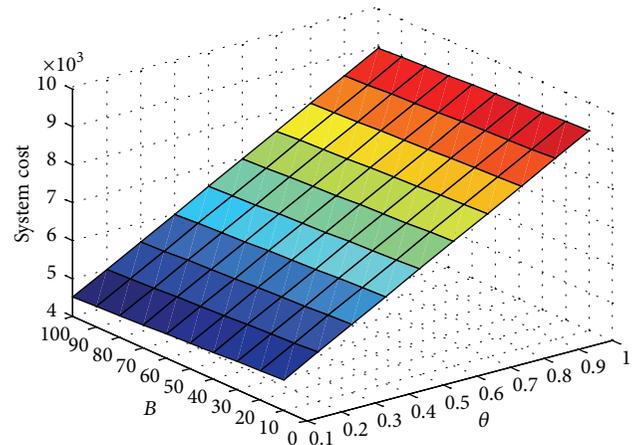
$\theta$	SODNDP	SODNDP solution time (sec)	SODNDP-RCI	SODNDP-RCI solution time (sec)	SODNDP-AARCI	SODNDP-AARCI solution time (sec)
0.1	4017.0113	0.08	16434.0362	0.06	4505.6073	83.09
0.2			18757.4762	0.06	5001.6873	34.63
0.3			21123.6929	0.06	5497.7673	65.28
0.4			23525.3110	0.06	5993.8473	93.75
0.5			26690.3091	0.06	6489.9273	86.87
0.6			30183.6424	0.05	6986.0073	36.84
0.7			33688.2109	0.06	7482.0873	50.98
0.8			37206.0000	0.05	7978.1673	83.61
0.9			40735.8000	0.06	8474.2473	48.67
1.0			44265.6000	0.03	8970.3273	188.84

TABLE 4: The total travel cost and runtimes of the SODNDP, SODNDP-RC1, and SODNDP-AARC1 models under different available budgets when  $\theta = 0.5$ .

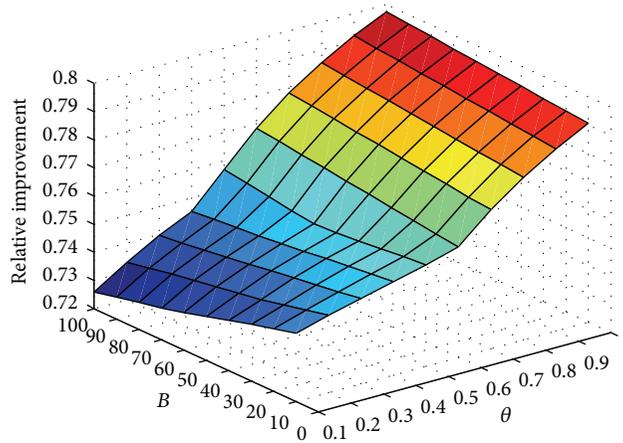
$B$	SODNDP	SODNDP solution time (sec)	SODNDP-RC1	SODNDP-RC1 solution time (sec)	SODNDP-AARC1	SODNDP-AARC1 solution time (sec)
10	4304.0727	0.05	27806.2868	0.06	6784.4727	25.30
20	4271.3455	0.06	27575.3434	0.06	6751.7455	34.95
30	4238.6182	0.06	27344.4000	0.06	6719.0182	29.41
40	4205.8909	0.06	27113.4566	0.06	6686.2909	67.55
50	4173.1636	0.08	26914.0258	0.05	6653.5636	26.08
60	4140.4364	0.06	26716.6065	0.05	6620.8364	32.89
70	4107.7091	0.06	26706.6727	0.06	6588.1091	59.81
80	4074.9818	0.05	26701.2182	0.05	6555.3818	68.64
90	4044.1811	0.06	26695.7636	0.06	6522.6545	79.20



(a) The optimal travel cost of the SODNDP-RC1



(b) The optimal travel cost of the SODNDP-AARC1



(c) The relative improvement of the optimal objective value of the SODNDP-RC1 against SODNDP-AARC1

FIGURE 2: The optimal objective value of the SODNDP-RC1 and SODNDP-AARC1 models, and the relative improvement of the optimal objective value of the SODNDP-RC1 against SODNDP-AARC1 under different available budgets and uncertainty levels.

TABLE 5: Comparison of simulation results.

B	Mean			Standard deviation			Maximum		
	SODNDP	SODNDP-RCI	SODNDP-AARCI	SODNDP	SODNDP-RCI	SODNDP-AARCI	SODNDP	SODNDP-RCI	SODNDP-AARCI
$\theta = 0.1$									
10	4610.0	4610.0	4610.0	1367.1	1367.1	1367.1	7614.0	7614.0	7614.0
20	4575.5	4575.5	4575.5	1362.2	1362.2	1362.2	7581.3	7581.3	7581.3
30	4551.3	4551.3	4551.3	1357.5	1357.5	1357.5	7548.6	7548.6	7548.6
40	4527.7	4527.7	4527.7	1352.1	1352.1	1352.1	7515.8	7515.8	7515.8
50	4505.0	4505.0	4505.0	1346.9	1346.9	1346.9	7483.1	7483.1	7483.1
60	4483.4	4483.4	4483.4	1340.8	1340.8	1340.8	7450.3	7450.3	7450.3
70	4463.1	4463.1	4463.1	1338.3	1338.3	1334.1	7421.4	7417.6	7417.6
80	4443.5	4431.7	4443.5	1337.7	1327.5	1327.5	7394.2	7384.9	7384.9
90	4421.8	4405.3	4424.6	1322.9	1336.5	1320.7	7367.1	7354.1	7352.2
100	4395.2	4378.9	4406.7	1322.1	1335.1	1313.7	7339.9	7326.9	7314.9
$\theta = 0.2$									
10	4610.0	4610.0	4610.0	1367.1	1367.1	1367.1	7614.0	7614.0	7614.0
20	4575.5	4575.5	4575.5	1362.2	1362.2	1362.2	7581.3	7581.3	7581.3
30	4551.3	4551.3	4551.3	1357.5	1357.5	1357.5	7548.6	7548.6	7548.6
40	4527.7	4527.7	4527.7	1352.2	1352.2	1352.1	7515.9	7515.8	7515.8
50	4500.8	4500.8	4505.0	1351.7	1346.9	1346.9	7488.7	7488.1	7483.1
60	4473.9	4473.9	4483.4	1351.3	1340.8	1340.8	7461.5	7461.5	7450.3
70	4447.3	4447.3	4463.1	1350.3	1334.1	1334.1	7434.4	7434.4	7417.6
80	4420.8	4420.8	4443.5	1349.1	1349.1	1327.5	7407.2	7407.2	7384.9
90	4394.7	4394.7	4424.6	1347.3	1347.3	1320.7	7380.0	7380.0	7352.2
100	4368.5	4368.5	4406.7	1345.5	1345.5	1313.7	7352.9	7352.9	7314.9
$\theta = 0.3$									
10	4610.0	4610.0	4610.0	1367.1	1367.1	1367.1	7614.0	7614.0	7614.0
20	4575.5	4575.5	4575.5	1363.7	1363.7	1362.2	7583.2	7583.2	7581.3
30	4548.1	4548.1	4551.3	1363.3	1363.3	1357.5	7556.0	7556.0	7548.6
40	4521.3	4521.3	4527.7	1362.6	1362.6	1352.1	7528.8	7528.8	7515.8
50	4494.7	4494.7	4505.0	1361.7	1361.7	1346.9	7501.7	7501.7	7483.1
60	4468.2	4468.2	4483.4	1360.5	1360.5	1340.8	7474.5	7474.5	7450.3
70	4441.7	4441.7	4463.1	1359.1	1359.1	1334.1	7447.3	7447.3	7417.6
80	4415.3	4415.3	4443.5	1357.7	1357.7	1327.5	7420.1	7420.1	7384.9
90	4390.2	4390.2	4424.6	1356.1	1356.1	1320.7	7394.1	7394.1	7352.2
100	4368.0	4368.0	4406.7	1354.4	1354.4	1313.7	7370.9	7370.9	7314.9

an increasing value of the uncertainty level  $\theta$ . However, the total travel cost from the adjustable robust solution is far lower than the robust solution under any circumstances. This also shows that the adjustable solution is less conservative than the robust solution. When the uncertainty level  $\theta \geq 0.7$ , the total travel cost from the robust solution basically remains unchanged with the increase of available budget  $B$ , which means that the increase of available budget has no much effect on the improvement of traffic congestion. This implies that some investments are wasted, and this case does not appear in the SODNDP-AARC1 model. Moreover, the relative improvement of the total travel cost of the robust solution against adjustable solution is monotonically decreasing function with respect to the available budget  $B$  and is monotonically increasing function with respect to the uncertainty level  $\theta$ . When the uncertainty level  $\theta \geq 0.5$ , the relative improvement range of the total travel cost of the robust solution against adjustable solution becomes larger, and it almost reaches 80% when the uncertainty level  $\theta$  equals 1. Thus, the SODNDP-AARC1 model is less conservative than the SODNDP-RC1 model; therefore, it becomes more attractive.

To compare the operating behaviors of the three capacity expansion plans, nominal, robust, and adjustable robust, we randomly generated 100 travel demand vectors in which the demand of each O-D pair satisfies the normal distribution whose mean is  $\bar{d}_i^t$  and standard deviation is  $0.5\bar{d}_i^t$ . For each random vector  $d_i^t$ , for all  $i \in C_R$ , the total travel cost from the nominal, robust, and adjustable robust capacity expansion plans is computed. The mean, standard deviation, and maximum of the total travel cost associated with nominal, robust, and adjustable robust capacity expansion plans are shown in Table 5. It can be seen that the mean, standard deviation, and maximum of the total travel cost from the nominal, robust, and adjustable robust capacity expansion plans all decrease as the available budget  $B$  increases. The mean and maximum of the total travel cost from the robust capacity expansion plans reduce when the uncertainty level  $\theta$  increases, and standard deviation of the total travel cost from the robust capacity expansion plans increases when the uncertainty level  $\theta$  increases. On the other hand, the mean, standard deviation, and maximum of the total travel cost from adjustable robust capacity expansion plans remain unchanged for all uncertainty levels. This is because the adjustable robust capacity expansion plans are the same under different uncertainty levels. In addition, the mean of the total travel cost from the adjustable robust capacity expansion plans is greater than that of the robust capacity expansion plans; however, in all case, the standard deviation and maximum of the total travel cost from the adjustable robust capacity expansion plans are less than or equal to those of the robust capacity expansion plans.

The optimal objective value of the SODNDP-RC1 and SODNDP-AARC1 models is denoted as  $z^*(b^R)$  and  $z^*(b^{AR})$ , respectively. Moreover, the optimal objective value of the SODNDP model is denoted as  $z^*(b^N)$ , where it is computed based on the nominal demand  $\bar{d}_i^t$ . Among the 100 travel demand vectors ( $d_i^t$ ), we compare total cost from

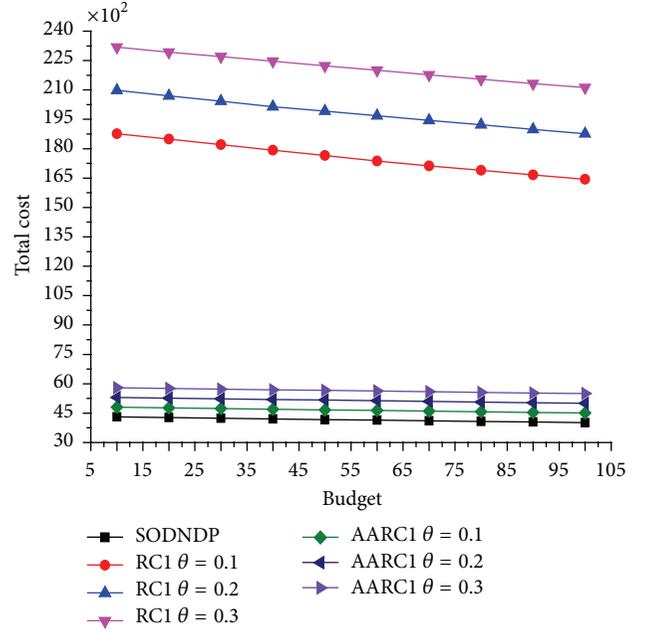


FIGURE 3: The optimal objective value of the SODNDP, SODNDP-RC1, and SODNDP-AARC1 models.

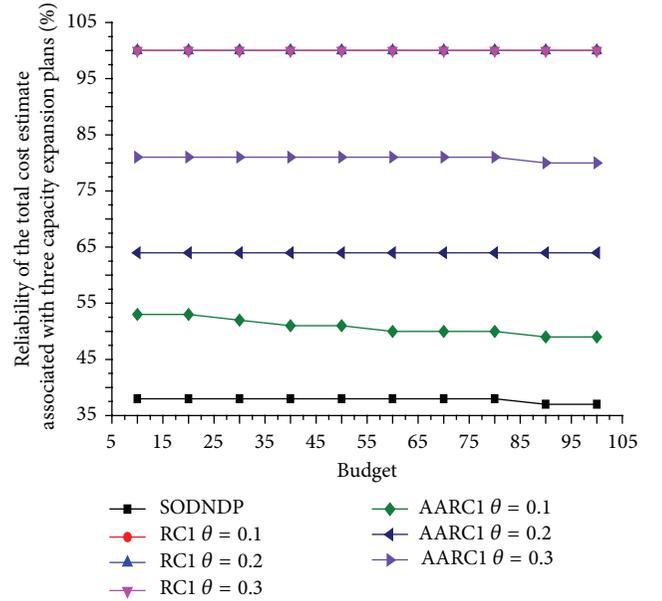


FIGURE 4: Reliability of the total cost estimate associated with nominal, robust, and adjustable robust capacity expansion plans.

the nominal, robust, and adjustable robust capacity expansion plans against  $z^*(b^N)$ ,  $z^*(b^R)$ , and  $z^*(b^{AR})$ , respectively. Figure 3 depicts the optimal objective value of the SODNDP, SODNDP-RC1, and SODNDP-AARC1 models,  $z^*(b^N)$ ,  $z^*(b^R)$ , and  $z^*(b^{AR})$ , under different available budgets and uncertainty levels. Figure 4 shows the percentages for which  $\sum_{t \in \mathfrak{S}, i \in C \setminus C_s} c_i^t x_i^t(d, b^N) \leq z^*(b^N)$ ,  $\sum_{t \in \mathfrak{S}, i \in C \setminus C_s} c_i^t x_i^t(d, b^R) \leq z^*(b^R)$  and  $\sum_{t \in \mathfrak{S}, i \in C \setminus C_s} c_i^t x_i^t(d, b^{AR}) \leq z^*(b^{AR})$  under different

available budgets and uncertainty levels. As shown in Figures 3 and 4, on average, approximately 38% or 38 random demands out of the 100 yields total travel cost no larger than  $z^*(b^N)$ . In other words, the nominal total travel cost is 38% reliable. On the other hand, the average reliability of the robust total travel cost is 100% under different uncertainty levels, and the average reliability of the adjustable robust total travel cost is 50% when uncertainty level  $\theta = 0.1$ . Although the average reliability of the robust total travel cost is higher than that of the adjustable total travel cost, the robust total travel cost is far greater than the adjustable robust total travel cost, which demonstrates that the robust capacity expansion plans are overly conservative. In addition, the average reliability of the adjustable robust total travel cost is monotonically increasing with the rise of uncertainty level  $\theta$ . The average reliability of the adjustable robust total travel cost is 80% when uncertainty level  $\theta = 0.3$ . The average reliability of the adjustable robust total travel cost is slightly lower than that of the robust total travel cost; however, the adjustable robust capacity expansion plans yield lower total travel cost. Thus, it is continuing to suggest that the adjustable robust capacity expansion plans are more flexible.

## 7. Conclusion and Further Research

This paper applied the adjustable robust optimization methodology to the CTM-based single-level system-optimal dynamic NDP (DNDP) under demand uncertainty. We assumed that the uncertain demand belongs to a polyhedral set and used the adjustable robust optimization methodology to formulate the AARC of DNDP. In addition, the RC of DNDP was formulated by using the usual robust optimization approach, which was compared against the AARC of DNDP. Numerical example results show that the adjustable robust capacity expansion plans are more flexible and significantly less conservative than the usual robust capacity expansion plans.

Several directions for future research are worth noting. Firstly, in this paper, the used LP based CTM model allows vehicle holding [35], which may be unrealistic. Secondly, ARO approach can be applied to alternative deterministic mathematical formulation [38] to overcome this problem. In addition, an extension to consider unbounded uncertainty set with globalized robust optimization [39] is another interesting research direction.

## Notations

### Sets

- $\mathfrak{T}$ : Set of time intervals  $\{1, 2, \dots, T\}$
- $C$ : Set of cells
- $C_R$ : Set of source cells
- $C_S$ : Set of sink cells
- $A$ : Adjacent matrix,  $A = \{a_{ij}\}$ ; if cell  $i$  is connected to cell  $j$ , then  $a_{ij} = 1$ ; otherwise  $a_{ij} = 0$ .

### Parameters

- $d_i^t$ : Demand generated in cell  $i$  at time  $t$ ,  $i \in C_R$
- $c_i^t$ : Travel cost in cell  $i$  at time  $t$
- $N_i^t$ : Capacity in cell  $i$  at time  $t$
- $\delta_i^t$ : Ratio of the free-flow speed over the back propagation speed of cell  $i$  at time  $t$
- $Q_i^t$ : Inflow/outflow capacity of cell  $i$  at time  $t$
- $B$ : Total investment budget
- $\chi_i$ : Increase in  $N_i^t$  for a unit increase of  $b_i$
- $\phi_i$ : Increase in  $Q_i^t$  for a unit increase of  $b_i$
- $\hat{x}_i$ : Initial number of vehicles of cell  $i$ .

### Variables

- $b_i$ : Investment cost spent on cell  $i$ ,  $b = (\dots, b_i, \dots)$
- $x_i^t$ : Number of vehicles staying in cell  $i$  at time  $t$ ,  $x = (\dots, x_i^t, \dots)$
- $y_{ij}^t$ : Number of vehicles moving from cell  $i$  to cell  $j$  at time  $t$ ,  $y = (\dots, y_{ij}^t, \dots)$ .

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The work described in this paper was jointly supported by grants from the National Natural Science Foundation of China (71131001) and the National Basic Research Program of China (2012CB725400).

## References

- [1] B. D. Chung, T. Yao, C. Xie, and A. Thorsen, "Robust optimization model for a dynamic design problem under demand uncertainty," *Networks and Spatial Economics*, vol. 11, no. 2, pp. 371–389, 2011.
- [2] S. T. Waller, J. L. Schofer, and A. K. Ziliaskopoulos, "Evaluation with traffic assignment under demand uncertainty," *Transportation Research Record*, no. 1771, pp. 69–74, 2001.
- [3] A. Chen, Z. Zhou, P. Chootinan, S. Ryu, C. Yang, and S. C. Wong, "Transportation network design under uncertainty: a review and new developments," *Transport Reviews*, vol. 31, no. 6, pp. 743–768, 2011.
- [4] T. L. Magnanti and R. T. Wong, "Network design and transportation planning: models and algorithms," *Transportation Science*, vol. 18, no. 1, pp. 1–55, 1984.
- [5] M. Minoux, "Network synthesis and optimum network design problems: models, solution methods and applications," *Networks*, vol. 19, no. 3, pp. 313–360, 1989.
- [6] H. Yang and M. G. H. Bell, "Models and algorithms for road network design: a review and some new developments," *Transport Reviews*, vol. 18, no. 3, pp. 257–278, 1998.
- [7] D.-Y. Lin, A. Karoonsoontawong, and S. T. Waller, "A Dantzig-Wolfe decomposition based heuristic scheme for bi-level

- dynamic network design problem,” *Networks and Spatial Economics*, vol. 11, no. 1, pp. 101–126, 2011.
- [8] K. Jeon, S. V. Ukkusuri, and S. T. Waller, “Heuristic approach for discrete network design problem accounting for dynamic traffic assignment conditions: formulation, solution methodologies, implementations and computation experiences,” in *Proceedings of the 84th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2005.
- [9] S. T. Waller, K. C. Mouskos, D. Kamaryannis, and A. K. Ziliaskopoulos, “A linear model for the continuous network design problem,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 21, no. 5, pp. 334–345, 2006.
- [10] S. V. Ukkusuri and S. T. Waller, “Linear programming models for the user and system optimal dynamic network design problem: formulations, comparisons and extensions,” *Networks and Spatial Economics*, vol. 8, no. 4, pp. 383–406, 2008.
- [11] D.-Y. Lin, “A dual variable approximation-based descent method for a bi-level continuous dynamic network design problem,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 26, no. 8, pp. 581–594, 2011.
- [12] B. N. Janson, “Network design effects of dynamic traffic assignment,” *Journal of Transportation Engineering*, vol. 121, no. 1, pp. 1–13, 1995.
- [13] S. T. Waller, *Optimization and control of stochastic dynamic transportation system: formulations, solution methodologies, and computational experience [Ph.D. dissertation]*, Northwestern University, 2000.
- [14] S. T. Waller and A. K. Ziliaskopoulos, “Stochastic dynamic network design problem,” *Transportation Research Record*, no. 1771, pp. 106–113, 2001.
- [15] A. Karoonsoontawong and S. T. Waller, “Comparison of system- and user-optimal stochastic dynamic network design models using Monte Carlo bounding techniques,” *Transportation Research Record*, vol. 1923, pp. 91–102, 2005.
- [16] A. Karoonsoontawong and S. T. Waller, “Dynamic continuous network design problem: linear bilevel programming and meta-heuristic approaches,” *Transportation Research Record*, no. 1964, pp. 104–117, 2006.
- [17] A. Karoonsoontawong and S. T. Waller, “Robust dynamic continuous network design problem,” *Transportation Research Record*, no. 2029, pp. 58–71, 2007.
- [18] A. Karoonsoontawong and S. T. Waller, “Integrated network capacity expansion and traffic signal optimization problem: robust bi-level dynamic formulation,” *Networks and Spatial Economics*, vol. 10, no. 4, pp. 525–550, 2010.
- [19] C. F. Daganzo, “The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory,” *Transportation Research Part B*, vol. 28, no. 4, pp. 269–287, 1994.
- [20] C. F. Daganzo, “The cell transmission model—part II: network traffic,” *Transportation Research Part B*, vol. 29, no. 2, pp. 79–93, 1995.
- [21] A. K. Ziliaskopoulos, “Linear programming model for the single destination System Optimum Dynamic Traffic Assignment problem,” *Transportation Science*, vol. 34, no. 1, pp. 37–49, 2000.
- [22] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski, “Adjustable robust solutions of uncertain linear programs,” *Mathematical Programming. A Publication of the Mathematical Programming Society*, vol. 99, no. 2, pp. 351–376, 2004.
- [23] A. Ben-Tal and A. Nemirovski, “Robust convex optimization,” *Mathematics of Operations Research*, vol. 23, no. 4, pp. 769–805, 1998.
- [24] A. Ben-Tal and A. Nemirovski, “Robust solutions of uncertain linear programs,” *Operations Research Letters*, vol. 25, no. 1, pp. 1–13, 1999.
- [25] A. Ben-Tal and A. Nemirovski, “Robust solutions of linear programming problems contaminated with uncertain data,” *Mathematical Programming*, vol. 88, no. 3, pp. 411–424, 2000.
- [26] A. Ben-Tal and A. Nemirovski, “Robust optimization—methodology and applications,” *Mathematical Programming*, vol. 92, no. 3, pp. 453–480, 2002.
- [27] D. Bertsimas and M. Sim, “The price of robustness,” *Operations Research*, vol. 52, no. 1, pp. 35–53, 2004.
- [28] J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios, “Robust optimization of large-scale systems,” *Operations Research*, vol. 43, no. 2, pp. 264–281, 1995.
- [29] Y. Yin and S. Lawphongpanich, “A robust approach to continuous network design problems with demand uncertainty,” in *Proceedings of 17th International Symposium of Transportation and Traffic Theory*, R. E. Allsop, M. G. H. Bell, and B. G. Heydecker, Eds., pp. 111–126, 2007.
- [30] Y. Yin, S. M. Madanat, and X.-Y. Lu, “Robust improvement schemes for road networks under demand uncertainty,” *European Journal of Operational Research*, vol. 198, no. 2, pp. 470–479, 2009.
- [31] Y. Lou, Y. Yin, and S. Lawphongpanich, “Robust approach to discrete network designs with demand uncertainty,” *Transportation Research Record*, no. 2090, pp. 86–94, 2009.
- [32] T. Yao, S. R. Mandala, and B. D. Chung, “Evacuation transportation planning under uncertainty: a robust optimization approach,” *Networks and Spatial Economics*, vol. 9, no. 2, pp. 171–189, 2009.
- [33] B. D. Chung, T. Yao, T. L. Friesz, and H. Liu, “Dynamic congestion pricing with demand uncertainty: a robust optimization approach,” *Transportation Research Part B*, vol. 46, no. 10, pp. 1504–1518, 2012.
- [34] A. Ben-Tal, B. D. Chung, S. R. Mandala, and T. Yao, “Robust optimization for emergency logistics planning: risk mitigation in humanitarian relief supply chains,” *Transportation Research Part B*, vol. 45, no. 8, pp. 1177–1189, 2011.
- [35] K. Doan and S. V. Ukkusuri, “On the holding back problem in the cell transmission based dynamic traffic assignment models,” *Transportation Research Part B*, vol. 46, no. 9, pp. 1218–1238.
- [36] S. Travis Waller and A. K. Ziliaskopoulos, “A chance-constrained based stochastic dynamic traffic assignment model: analysis, formulation and solution algorithms,” *Transportation Research C*, vol. 14, no. 6, pp. 418–427, 2006.
- [37] A. Brooke, D. Kendrick, and A. Meeraus, *GAMS: A User’s Guide*, The Scientific Press, San Francisco, Calif, USA, 1992.
- [38] F. Zhu and S. V. Ukkusuri, “A cell based dynamic system optimum model with non-holding back flows,” *Transportation Research Part C*, vol. 36, pp. 367–380, 2013.
- [39] A. Ben-Tal, S. Boyd, and A. Nemirovski, “Extending scope of robust optimization: comprehensive robust counterparts of uncertain problems,” *Mathematical Programming*, vol. 107, no. 1–2, pp. 63–89, 2006.
- [40] Y. Li, S. T. Waller, and T. Ziliaskopoulos, “A decomposition scheme for system optimal dynamic traffic assignment models,” *Network Spatial Economic*, vol. 3, no. 4, pp. 441–455, 2003.

## Research Article

# Congestion Behavior under Uncertainty on Morning Commute with Preferred Arrival Time Interval

LingLing Xiao,<sup>1</sup> Ronghui Liu,<sup>2</sup> and HaiJun Huang<sup>1</sup>

<sup>1</sup> School of Economics and Management, Beihang University, Beijing 100191, China

<sup>2</sup> Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK

Correspondence should be addressed to Ronghui Liu; r.liu@its.leeds.ac.uk

Received 5 November 2013; Accepted 18 December 2013; Published 17 February 2014

Academic Editor: Huimin Niu

Copyright © 2014 LingLing Xiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper extends the bottleneck model to study congestion behavior of morning commute with flexible work schedule. The proposed model assumes a stochastic bottleneck capacity which follows a uniform distribution and homogeneous commuters who have the same preferred arrival time interval. The commuters are fully aware of the stochastic properties of travel time and schedule delay distributions at all departure times that emerge from day-to-day capacity variations. The commuters' departure time choice follows user equilibrium (UE) principle in terms of the expected trip cost. Analytical and numerical solutions of this model are provided. The equilibrium departure time patterns are examined which show that the stochastic capacity increases the mean trip cost and lengthens the rush hour. The adoption of flexitime results in less congestion and more efficient use of bottleneck capacity than fixed-time work schedule. The longer the flexi-time interval is, the more uniformly distributed the departure times are.

## 1. Introduction

The bottleneck model was first proposed by Vickrey [1] and has subsequently inspired many to develop more realistic extensions, to gain qualitative and theoretical insights into transport policy measures on commuter travel behavior (e.g., Smith [2]; Daganzo [3]; Braid [4]; Arnott et al. [5]; Yang and Huang [6]; Ramadurai et al. [7]). In these analyses, commuters must choose their departure times to minimize the sum of travel delays (depending on the bottleneck capacity) and schedule delays (depending on the difference between actual and desired arrival times). At user equilibrium, no commuter could reduce the total commute cost by unilaterally changing his/her departure time.

The study of decision making under risk (and uncertainty) has a long history in the fields of economics, psychology, transport, and beyond (Machina [8]; Cubitt and Sugden [9]; Gollier and Treich [10]; Birnbaum and Schmidt [11]). Many mathematical tools and analytical frameworks are used to model people's behavior and predict likely choice outcomes in varying settings. The theory has also evolved from expected utility to nonexpected utility and been used to describe choice behavior in risky situations (Starmer [12]; de Palma et al. [13]).

This paper applies the expected utility theories to capture the departure time choice behavior in morning commute problem under uncertainty. Morning commute plays an important role in a monocentric city and the traffic congestion in such network is caused by concentration of travel demand around the work start time. The introduction of flexible work schedule is one of the transport demand management measures for alleviating peak congestion. The paper provides useful insight into traveler's decision making in contrast to fixed-time schedule pattern. Henderson [14] incorporated the productivity effect to analyze the equilibrium and optimum solution with staggered work hours. Mun and Yonekawa [15] extended bottleneck congestion to study the case that part of the firms in city adopts flexitime and incorporated the effects on urban productivity.

Most literatures on morning commute have assumed that the capacity of the bottleneck is deterministic and the traffic demand is also deterministic or governed by a predetermined elastic demand function. In reality, not only does travel time increase as traffic volume increases towards capacity but also the travel time becomes increasingly random and unpredictable due to the chaotic behavior of traffic at the micro level. The source of variation in road capacity may

occur due to physical and operational factors, such as road repairs, construction, accidents, and bad weather. The variations in road capacity from physical and operational reasons are what make the analysis of travel behavior so complex and yet interesting. As such, understanding travelers' attitudes and their behavior in varying settings is key to developing sustainable transport policies. There has been recent attention to the stochastic nature of the bottleneck models (Siu and Lo [16]; Li et al. [17]; Xiao et al. [18]). A reasonable way to capture these variations and their impact on the network performance is to formulate the problem using probability distributions (Chen et al. [19]).

The focus of this paper is to analyze the departure time choice behavior under uncertainty in the morning commuting problem with flexi-time work schedule. It is expected that the stochastic capacity leads to uncertainty in queuing, travel time, and trip cost, which in turn influences the commuters' travel choice behavior. We assume that travelers are fully aware of the stochastic properties of the travel time and schedule delay distributions throughout the morning peak period which emerges from their day-to-day travel experience. Furthermore, we consider homogenous travelers have the same preferred arrival time interval (PATI). We formulate a stochastic bottleneck model for this flexi-time commute problem and derive its analytical solution. The properties of the model are investigated.

The solution of the proposed model shows that the capacity variability of the bottleneck leads to significant changes in departure time patterns, which are different to those derived under deterministic conditions. In a deterministic bottleneck model with flexi-time work schedule, an individual can choose either to depart in the tails of the rush hour when travel time is low and pay the penalty of arriving at work early or late, or to depart close to the PATI when travel time is high but schedule delay cost is low. In other words, under the deterministic equilibrium, schedule delay early/late and arrival on time cannot occur simultaneously for a given departure time (Henderson [14]; Mun and Yonekawa [15]). We demonstrate that, with day-to-day stochastic capacity, commuters departing at the same time may endure schedule delay early/late or not and may experience queuing delay or not on different days.

The rest of this paper is organized as follows. Section 2 formulates separately a deterministic and a stochastic bottleneck model with homogeneous PATI. The equilibrium solutions for the departure time pattern are derived for each case. The theoretical properties of the proposed stochastic bottleneck model with PATI are investigated and compared with the deterministic case. Numerical examples are presented in Section 3 to illustrate further the equilibrium properties of the model. Section 4 provides conclusion remarks.

## 2. Departure Patterns of Morning Commute with Flexible Arrival Time

*2.1. The Deterministic Case.* We formulate the peak period congestion based on the bottleneck model developed by Vickrey [1]. Suppose a single road connecting a residential

area and the Commercial Business District (CBD), which has a bottleneck just before the CBD. It is assumed that vehicles drive at constant speed from home to the bottleneck point: travel time for this portion of trip is constant and represented as  $T_{\text{free}}$ . Queue develops when traffic flow rate exceeds the bottleneck capacity  $s$ . Travel time for a vehicle departing at time instant  $t$  is represented as follows:

$$T(t) = T_{\text{free}} + \frac{Q(t)}{s}, \quad (1)$$

where  $T(t)$  is travel time and  $Q(t)$  is length of queue. The second term represents the waiting time within the queue behind the bottleneck. We set  $T_{\text{free}} = 0$ , hereafter. This setting will not affect the qualitative property.

The queue length that a trip maker departing at time  $t$  encounters is calculated as follows:

$$Q(t) = \max \left\{ \int_{t_0}^t [r(x) - s] dx, 0 \right\}, \quad (2)$$

and the cumulative departures as,

$$R(t) = \int_{t_0}^t r(x) dx, \quad (3)$$

where  $r(x)$  is the departure rate at time instant  $t$  and  $t_0$  is the earliest time with positive departure rate.

Suppose that every morning, a fixed number of  $N$  individuals commute from home to office located in the CBD, driving along the road stated above. All workers have identical skills and preferences. Unlike Vickrey [1] which assumes that there is only one preferred arrive time  $t^*$ , here, firms in the CBD adopt a flexi-time work schedule such that employees arriving at office earlier or later  $\delta$  ( $\delta \geq 0$ ) than  $t^*$  incur no scheduling cost. Hereafter, we call this time period,  $[t^* - \delta, t^* + \delta]$ , as PATI.

Some commuters may still arrive at the destination earlier or later than PATI, in order to avoid a long queue at the bottleneck. The cost for commuters traveling from home to the CBD consists of three components: the cost of travel time and the cost of schedule delay early or late. It can be formulated as follows:

$$C(t) = \alpha T(t) + \text{SDE}(t) + \text{SDL}(t), \quad (4)$$

where  $\alpha$  is the value of travel time. The cost of schedule delay early (SDE) and schedule delay late (SDL) for a commuter who leaves home at time  $t$  can be expressed as

$$\begin{aligned} \text{SDE}(t) &= \beta (t^* - \delta - (t + T(t))), \\ \text{SDL}(t) &= \gamma (t + T(t) - (t^* + \delta)), \end{aligned} \quad (5)$$

where  $\beta$  and  $\gamma$  denote the value of schedule delay early and the value of schedule delay late, respectively.

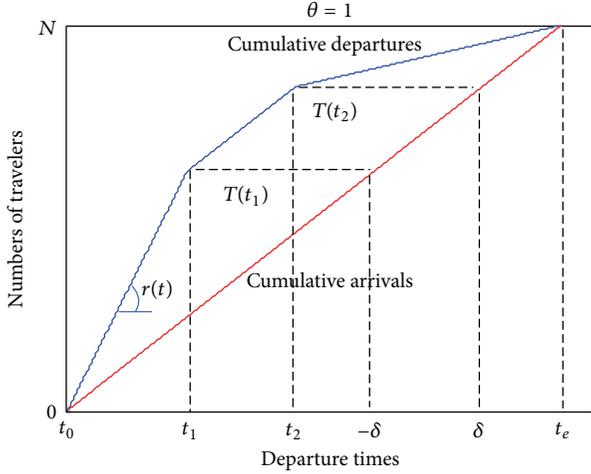


FIGURE 1: Departure time distributions in the deterministic case.

Substituting (1) and (5) into (4) and applying the equilibrium condition,  $\partial C(t)/\partial t = 0$ , we have the equilibrium departure rate as follows,

$$r(t) = \begin{cases} \frac{\alpha s}{(\alpha - \beta)}, & \text{if } t_0 \leq t < t_1 \\ s, & \text{if } t_1 \leq t \leq t_2 \\ \frac{\alpha s}{(\alpha + \gamma)}, & \text{if } t_2 < t \leq t_e, \end{cases} \quad (6)$$

where  $t_0$  and  $t_e$  are the earliest time and the latest time with positive departure rate, respectively. And  $t_1$  and  $t_2$  are the watershed times for which an individual arrives at work on time. Then

$$\begin{aligned} t_0 &= t^* + \frac{\gamma - \beta}{\beta + \gamma} \delta - \frac{\gamma}{\beta + \gamma} \cdot \frac{N}{s}, \\ t_1 &= \frac{\alpha - \beta}{\alpha} t^* - \frac{\alpha - \beta}{\alpha} \delta - \frac{\beta \gamma}{\alpha(\beta + \gamma)} \cdot \frac{N}{s}, \\ t_2 &= \frac{\alpha - \beta}{\alpha} t^* + \frac{\alpha + \beta}{\alpha} \delta - \frac{\beta \gamma}{\alpha(\beta + \gamma)} \cdot \frac{N}{s}, \\ t_e &= t^* + \frac{\gamma - \beta}{\beta + \gamma} \delta + \frac{\beta}{\beta + \gamma} \cdot \frac{N}{s}. \end{aligned} \quad (7)$$

Figure 1 depicts the cumulative departures (from home) and arrivals (at CBD) in equilibrium under the deterministic capacity case. For simplicity, we set  $t^*$  to be zero, and then all travelers have the same PATI  $[-\delta, \delta]$ . The horizontal distance between the departure and arrival curves gives the travel time. Several critical time points are indicated. Travelers departing at time interval  $[t_1, t_2]$  have the longest travel time but will arrive on time, that is, within PATI. Travelers departing before  $t_1$  will arrive earlier than desired, while the travelers departing after  $t_2$  will arrive late at the destination.

**2.2. The Stochastic Case.** The deterministic case models a single-day departure time equilibrium. In the real world, road

capacities may vary from day to day due to unexpected events such as incidents and weather conditions. Because of the capacity fluctuations, both commuters' travel time and their schedule delays are stochastic. In this section, we hypothesize that a constant long-term departure time pattern may emerge given the responses of the travelers to the day-to-day capacity variation. Each commuter chooses an optimal departure time which minimizes his/her long-term expected trip cost. We call this pattern, if it exists, a long-term equilibrium pattern.

**2.2.1. Assumptions and Travelers' Cost Function.** The following assumptions are made in the model formulation.

- (A1) Commuters are homogeneous with the same  $\alpha$ ,  $\beta$ , and  $\gamma$  values and the same PATI.
- (A2) The capacity of the bottleneck is constant within a day but fluctuates from day to day. The uncertainty of capacity is completely exogenous and independent of departures.
- (A3) The capacity is a nonnegative stochastic variable changing around a certain mean capacity. Following Li et al. [17], we assume that stochastic capacity follows a uniform distribution within interval  $[\theta \bar{s}, \bar{s}]$ , where  $\bar{s}$  is the design capacity and  $\theta (< 1)$  is a positive parameter which denotes the lowest rate of available capacity.
- (A4) Commuters are aware of the capacity degeneration probability, and their departure time choice follows the user equilibrium (UE) principle in terms of mean trip cost.

We assume that the capacity of the single bottleneck is stochastic but the commuters' departure time choice is deterministic. The calculation of the mean trip cost relies on the calculations of the mean travel time, the mean schedule delay early and late. For simplicity, we set the  $t^*$  to be zero, and then the PATI becomes  $[-\delta, \delta]$ . Under the stochastic condition, the mean trip cost with respect to departure time  $t$  can be formulated as follows:

$$E[c(t)] = E[\alpha T(t) + SDE(t) + SDL(t)]. \quad (8)$$

The equilibrium condition for commuters' departure time choice in a single bottleneck with stochastic capacity is that no commuter can reduce his/her mean trip cost by unilaterally altering his/her departure time. This condition implies that the commuters' mean trip cost is fixed with respect to the time instant with positive departure rate. That is

$$\frac{\partial E[c(t)]}{\partial t} = 0, \quad \text{if } r(t) > 0. \quad (9)$$

**2.2.2. Mathematical Formulations and Derivations.** Due to the stochastic capacity over days, the travel time experienced by a traveler departing at the same time  $t$  varies from day to day. This is equivalent to saying that commuters departing home to work at the same time may endure schedule delay early/late or not and may experience queuing delay or not on different days. Consequently, there are six situations which

may occur: (I) always arrive early, (II) possibly arrive early or on time, (III) always arrive on time, (IV) possibly arrive on time or late, (V) always arrive late, and (VI) always arrive late but queue may exist. We propose the following simple extension of trip cost function under the six situations to model user departure time choice under degradable capacities, and we use  $t_1, t_2, t_3, t_4$ , and  $t_5$  to denote the watershed lines separating the six cases.

As we assume that stochastic capacity is completely exogenous and independent of departure flows, the expectation of travel time and schedule delay cost with respect to different situations can be derived respectively as follows:

$$\begin{aligned}
E[T(t)] &= \begin{cases} \int_{\theta\bar{s}}^{\bar{s}} \left( \frac{R(t)}{s} + t_0 - t \right) f(s) ds, & t_0 \leq t \leq t_5 \\ \int_{\theta\bar{s}}^{R(t)/(t-t_0)} \left( \frac{R(t)}{s} - t + t_0 \right) f(s) ds, & t_5 < t \leq t_e \end{cases} \\
E[SDE(t)] &= \begin{cases} \beta \int_{\theta\bar{s}}^{\bar{s}} \left( -\delta - \frac{R(t)}{s} - t_0 \right) f(s) ds, & t_0 \leq t \leq t_1 \\ \beta \int_{R(t)/(-\delta-t_0)}^{\bar{s}} \left( -\delta - \frac{R(t)}{s} - t_0 \right) f(s) ds, & t_1 < t \leq t_2, \end{cases} \\
E[SDL(t)] &= \begin{cases} \gamma \int_{\theta\bar{s}}^{R(t)/(\delta-t_0)} \left( \frac{R(t)}{s} + t_0 - \delta \right) f(s) ds, & t_3 < t \leq t_4 \\ \gamma \int_{\theta\bar{s}}^{\bar{s}} \left( \frac{R(t)}{s} + t_0 - \delta \right) f(s) ds, & t_4 \leq t \leq t_5, \end{cases} \quad (10)
\end{aligned}$$

where  $f(s) = 1/(s - \theta s)$ . Substituting (10) into (8), we get the expected trip cost with respect to each situation. According to (9), the equilibrium departure rates for the six situations can be expressed as follows.

*Situation 1.* No commuters experience schedule delay later subject to all possible values of the bottleneck capacity. We get the departure rate

$$r(t) = \frac{\alpha}{\alpha - \beta} \cdot \frac{\bar{s}(1 - \theta)}{\ln \theta^{-1}}, \quad t_0 \leq t < t_1. \quad (11)$$

The boundary condition for this situation is  $SDL(t_1) = 0$  when  $s = \theta\bar{s}$ , and hence we have  $R(t_1) = -(t_0 + \delta)\theta\bar{s}$ .

*Situation 2.* If the capacity of the bottleneck is large enough, only schedule delay early will occur. On the contrary, no schedule delay occurs when the capacity is small. The watershed capacity satisfies  $T(t) + t = -\delta$ . Equivalently, we have  $s = R(t)/(-\delta - t_0)$  and the departure rate,

$$r(t) = \frac{\alpha\bar{s}(1 - \theta)}{\alpha \ln \theta^{-1} - \beta (\ln((-\delta - t_0)\bar{s}) - \ln R(t))}, \quad t_1 < t \leq t_2. \quad (12)$$

The boundary condition for this case is  $SDL(t_2) = 0$  when  $s = \bar{s}$ , and hence we have  $R(t_2) = -(t_0 + \delta)\bar{s}$ .

*Situation 3.* No commuters experience schedule delay subject to all possible values of the bottleneck capacity. Therefore, the departure rate is

$$r(t) = \frac{\bar{s}(1 - \theta)}{\ln \theta^{-1}}, \quad t_2 < t \leq t_3. \quad (13)$$

The boundary condition for this case is  $SDE(t_3) = SDL(t_3) = 0$  when  $s = \theta\bar{s}$ . Hence, we have  $R(t_3) = (\delta - t_0)\theta\bar{s}$ .

*Situation 4.* If the capacity of the bottleneck is large enough, individuals arrive on time. On the contrary, schedule delay late occurs when the capacity is small. The watershed capacity satisfies  $T(t) + t = \delta$ . Equivalently, we have  $s = R(t)/(\delta - t_0)$ . Therefore,

$$r(t) = \frac{\alpha\bar{s}(1 - \theta)}{\alpha \ln \theta^{-1} + \gamma (\ln R(t) - \ln((\delta - t_0)\theta\bar{s}))}, \quad t_3 < t \leq t_4. \quad (14)$$

The boundary condition for this case is  $SDE(t_4) = 0$  when  $s = \bar{s}$ , and then we have  $R(t_4) = (\delta - t_0)\bar{s}$ .

*Situation 5.* Similar to Situation 1, we have

$$r(t) = \frac{\alpha}{\alpha + \gamma} \cdot \frac{\bar{s}(1 - \theta)}{\ln \theta^{-1}}, \quad t_4 < t \leq t_5. \quad (15)$$

The boundary condition for this case is  $R(t_5) = \bar{s}(t_5 - t_0)$ ; that is, the queue length at time  $t_5$  equals to zero when  $s = \bar{s}$ .

*Situation 6.* Similar to Situation 2, we can find a watershed capacity of the bottleneck such that the queue length equals zero; that is,  $R(t) = s(t - t_0)$ , and hence the watershed capacity is  $R(t)/(t - t_0)$ . We have

$$r(t) = \frac{dR(t)}{dt} = \frac{(\alpha + \gamma)R(t)/(t - t_0) - (\alpha\theta + \gamma)\bar{s}}{(\alpha + \gamma)(\ln R(t) - \ln \theta\bar{s}(t - t_0))}, \quad t_3 < t \leq t_e. \quad (16)$$

The boundary condition for this case is  $r(t_e) = 0$ . Equivalently, we have  $R(t_e) = \hat{s}(t_e - t_0)$ , where  $\hat{s} = \bar{s}(\alpha\theta + \gamma)/(\alpha + \gamma)$  is the mean capacity under the stochastic case.

**2.2.3. Determination of the Watershed Time Instants.** Since the departure rate  $r(t) = 0$  if  $t > t_e$ , the cumulative departures at time  $t_e$  are equal to the traffic demand; that is,  $R(t_e) = N = \hat{s}(t_e - t_0)$ . Therefore, we have  $t_e = t_0 + N/\hat{s}$ . Moreover, the equilibrium condition of the stochastic bottleneck implies that  $E[c(t_0)] = E[c(t_e)]$ , and hence we have

$$t_0 = \frac{N}{\hat{s}} \cdot \frac{1}{\omega_0 - 1} + \frac{v_0}{1 - \omega_0} \delta, \quad t_e = \frac{N}{\hat{s}} \cdot \frac{\omega_0}{\omega_0 - 1} + \frac{v_0}{1 - \omega_0} \delta, \quad (17)$$

where

$$\begin{aligned}
\omega_0 &= 1 - \frac{(1 - \theta)(\beta + \gamma)}{(\alpha\theta + \gamma)(\ln \hat{s} - \ln \theta\bar{s})}, \\
v_0 &= \frac{(1 - \theta)(\gamma - \beta)}{(\alpha\theta + \gamma)(\ln \hat{s} - \ln \theta\bar{s})}.
\end{aligned} \quad (18)$$

Using the boundary conditions of Situations 1–6, we can obtain the watershed lines as follows:

$$\begin{aligned} t_1 &= \omega_1 t_0 + v_1 \delta, & t_2 &= \omega_2 t_0 + v_2 \delta, \\ t_3 &= \omega_3 t_0 + v_3 \delta, & t_4 &= \omega_4 t_0 + v_4 \delta, \\ t_5 &= \omega_5 t_0 + v_5 \delta, \end{aligned} \quad (19)$$

where

$$\begin{aligned} \omega_1 &= 1 - \frac{(\alpha - \beta)\theta\xi}{\alpha}, & \omega_2 &= \frac{(\alpha + \beta)}{\alpha} - \xi, \\ \omega_3 &= \frac{\beta}{\alpha} - \theta\xi + 1, & \omega_4 &= \frac{(\alpha + \gamma)(1 - \xi)}{\alpha} + \frac{\beta}{\alpha}, \\ \omega_5 &= 1 + \frac{(\gamma + \beta)}{(\alpha - (\alpha + \gamma)\xi)}, & v_1 &= \frac{(\beta - \alpha)\theta\xi}{\alpha}, \\ v_2 &= \frac{\beta}{\alpha} - \xi, & v_3 &= \frac{\beta}{\alpha} + \theta\xi, \\ v_4 &= \frac{((\alpha + \gamma)\xi + (\beta - \gamma))}{\alpha}, & v_5 &= \frac{(\beta - \gamma)}{(\alpha - (\alpha + \gamma)\xi)}, \\ \xi &= \frac{\ln \theta^{-1}}{(1 - \theta)}. \end{aligned} \quad (20)$$

With the resulting stochastic departure pattern at long-term equilibrium, the experienced day-to-day travel times and number of travelers experiencing queues change according to varied capacities over days. Given the boundary conditions of Situations 1–6, the cumulative departures and arrivals of a stochastic bottleneck are given in Figure 2. The two solid curves denote the cumulative departures and cumulative arrivals, as in Figure 1. The dotted lines are the maximum and the minimum capacity of the bottleneck, that is,  $\bar{s}$  and  $\theta\bar{s}$ . The earliest departure time from home to workplace is  $t_0$ , and if the bottleneck capacity equals  $\theta\bar{s}$ , commuters departing between  $t_1$  and  $t_3$  will arrive at workplace on time. If the capacity equals  $\bar{s}$ , commuters departing between  $t_2$  and  $t_4$  will arrive at workplace on time. Commuters departing within time interval  $[t_2, t_3]$  will always arrive at workplace on time, whatever the variability of capacity is. At the beginning, commuters depart from home with a constant departure rate until the watershed time  $t_1$ . Subsequently, the departure rate will gradually drop down. After time instant  $t_4$ , another constant departure rate will last until the watershed time instant  $t_5$  and the queuing length at this time will be zero. Later, the departure rate continues to drop down to zero at time  $t_e$ .

**2.2.4. Properties of the Stochastic Bottleneck Model.** In this subsection, we investigate the theoretical properties of the equilibrium solution of the proposed stochastic bottleneck model with PATI.

**Theorem 1.** *At equilibrium, the expected trip cost for all commuters is a monotonically increasing function of traffic demand and a monotonically decreasing function of  $\delta$ -value; that is,  $\partial E[c(t_0)]/\partial N > 0$  and  $\partial E[c(t_0)]/\partial \delta < 0$  hold.*

*Proof.* Since  $E[c(t_0)] = (-\delta - t_0)\beta$  and  $t_0 = N/(\bar{s}(\omega_0 - 1)) + v_0\delta/(1 - \omega_0)$ , we have  $\partial E[c(t_0)]/\partial \delta = -2\beta\gamma/(\beta + \gamma) < 0$  and

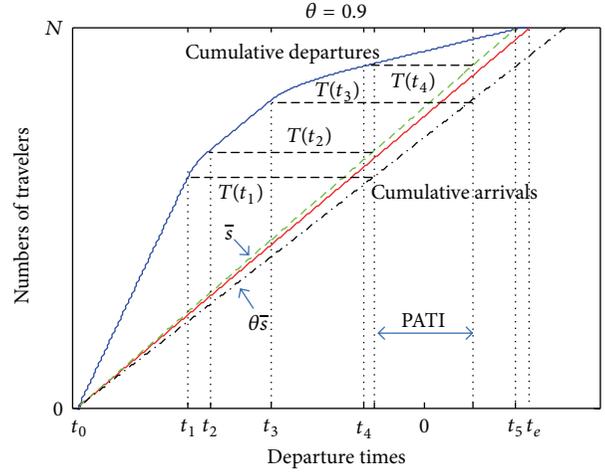


FIGURE 2: Departure time distributions in the stochastic case.

$\partial E[c(t_0)]/\partial N = \beta/((1 - \omega_0)\bar{s})$ . The definitions of  $\beta$  and  $\bar{s}$  imply that they are positive; hence, to prove  $\partial E[c(t_0)]/\partial N > 0$ , we only need to prove  $1 - \omega_0 > 0$ . Since  $0 < \theta < 1$  and  $\bar{s} = \bar{s}(\alpha\theta + \gamma)/(\alpha + \gamma) > \bar{s}(\alpha\theta + \gamma\theta)/(\alpha + \gamma) = \theta\bar{s}$ , we have  $1 - \theta > 0$ . And  $\ln \bar{s} - \ln \theta\bar{s} > 0$  clearly holds. Therefore, both the numerator and denominator of the second term in the right-hand side of the first equation of (18) are positive; therefore, we have  $1 - \omega_0 > 0$ . This completes the proof.  $\square$

**Theorem 2.** *At equilibrium, the expected trip cost for all commuters is a monotonically decreasing function of parameter  $\theta$ -value; that is,  $\partial E[c(t_0)]/\partial \theta < 0$  holds.*

*Proof.* Submitting  $t_0 = N/(\bar{s}(\omega_0 - 1)) + v_0\delta/(1 - \omega_0)$  into  $E[c(t_0)] = (-\delta - t_0)\beta$ , then the first-order derivative can be given as follows:

$$\frac{\partial E[c]}{\partial \theta} = \beta \frac{N}{\bar{s}} \cdot \frac{\alpha + \gamma}{\beta + \gamma} p'(\theta), \quad (21)$$

where

$$\begin{aligned} p(\theta) &= \frac{\ln \bar{s} - \ln(\theta\bar{s})}{1 - \theta}, \\ p'(\theta) &= \frac{1}{1 - \theta} \left( p(\theta) - \frac{\gamma}{\theta(\alpha\theta + \gamma)} \right). \end{aligned} \quad (22)$$

Since  $\ln(\bar{s}/(\theta\bar{s})) < 1 - \theta\bar{s}/\bar{s} = \gamma(1 - \theta)/(\alpha\theta + \gamma)$ , we then have

$$\begin{aligned} p(\theta) - \frac{\gamma}{\theta(\alpha\theta + \gamma)} & < \frac{\gamma}{\alpha\theta + \gamma} - \frac{\gamma}{\theta(\alpha\theta + \gamma)} = \frac{\gamma}{\alpha\theta + \gamma} \left( 1 - \frac{1}{\theta} \right) < 0. \end{aligned} \quad (23)$$

It is clear that  $p'(\theta) < 0$  and therefore we get  $\partial E[c]/\partial \theta < 0$ . This completes the proof.  $\square$

**Theorem 3.** *At equilibrium, the departure rate is a monotonically decreasing function of the departure time throughout the whole peak period; that is,  $dr(t)/dt \leq 0$ ,  $t \in [t_0, t_e]$ .*

*Proof.* According to (11)–(16), the departure rate  $r(t)$  is continuous within each of intervals  $[t_0, t_1)$ ,  $(t_1, t_2)$ ,  $(t_2, t_3)$ ,  $(t_3, t_4)$ ,  $(t_4, t_5)$ , and  $(t_5, t_e]$ . To prove the departure rate is continuous during the peak period, we calculate the following limitations:

$$\begin{aligned}
\lim_{t \rightarrow t_1^+} r(t) &= \lim_{t \rightarrow t_1^+} \frac{\alpha \bar{s}(1-\theta)}{\alpha \ln \theta^{-1} - \beta (\ln((-\delta - t_0) \bar{s}) - \ln R(t))} \\
&= \frac{\alpha}{\alpha - \beta} \cdot \frac{\bar{s}(1-\theta)}{\ln \theta^{-1}} = \lim_{t \rightarrow t_1^-} r(t), \\
\lim_{t \rightarrow t_2^-} r(t) &= \frac{\alpha \bar{s}(1-\theta)}{\alpha \ln \theta^{-1} - \beta (\ln((-\delta - t_0) \bar{s}) - \ln R(t))} \\
&= \frac{\bar{s}(1-\theta)}{\ln \theta^{-1}} = \lim_{t \rightarrow t_2^+} r(t), \\
\lim_{t \rightarrow t_3^-} r(t) &= \frac{\bar{s}(1-\theta)}{\ln \theta^{-1}} \\
&= \frac{\alpha \bar{s}(1-\theta)}{\alpha \ln \theta^{-1} + \gamma (\ln R(t) - \ln((\delta - t_0) \theta \bar{s}))} \\
&= \lim_{t \rightarrow t_3^+} r(t), \\
\lim_{t \rightarrow t_4^-} r(t) &= \frac{\alpha \bar{s}(1-\theta)}{\alpha \ln \theta^{-1} + \gamma (\ln R(t) - \ln((\delta - t_0) \theta \bar{s}))} \\
&= \frac{\alpha}{\alpha + \gamma} \cdot \frac{\bar{s}(1-\theta)}{\ln \theta^{-1}} = \lim_{t \rightarrow t_4^+} r(t), \\
\lim_{t \rightarrow t_5^+} r(t) &= \frac{(\alpha + \gamma) R(t_3) / (t_3 - t_0) - (\alpha \theta + \gamma) \bar{s}}{(\alpha + \gamma) \ln(R(t_3) / (\theta \bar{s}(t_3 - t_0)))} \\
&= \frac{\alpha}{\alpha + \gamma} \cdot \frac{\bar{s}(1-\theta)}{\ln \theta^{-1}} = \lim_{t \rightarrow t_5^-} r(t).
\end{aligned} \tag{24}$$

This proves  $r(t)$  is continuous indeed within the interval  $[t_0, t_e]$ .

Equations (11), (13), and (15) state that the departure rate  $r(t)$  is constant for  $t_0 \leq t \leq t_1$ ,  $t_2 \leq t \leq t_3$ , and  $t_4 \leq t \leq t_5$ , hence it is monotonically decreasing within these three intervals. By definition, the cumulative departure flow  $R(t)$  is nondecreasing with respect to time  $t$ . Thus, the denominators of the right-hand sides in (12) and (14) are nondecreasing with respect to time  $t$ . Therefore, the right-hand sides of (12) and (14) are nonincreasing with respect to time  $t$ ; that is, the departure rate  $r(t)$  is monotonically decreasing within  $[t_1, t_2]$  and  $[t_3, t_4]$ . The proof of  $dr(t)/dt \leq 0$  for all  $t \in (t_5, t_e]$  can be found in Xiao et al. [18].

In summary, the departure rate  $r(t)$  is monotonically decreasing within all four intervals and at their boundaries. Considering the continuity of  $r(t)$  for all  $t \in [t_0, t_e]$ , we conclude that  $r(t)$  is monotonically decreasing within  $[t_0, t_e]$ . This completes the proof.  $\square$

**Proposition 4.** *When parameter  $\theta$  approaches one, the stochastic bottleneck model immediately follows the deterministic model.*

*Proof.* According to the L'Hôpital's rule, we have  $\lim_{\theta \rightarrow 1} (1 - \theta) / (\ln \theta^{-1}) = 1$ . We then have

$$\begin{aligned}
\lim_{\theta \rightarrow 1} \omega_0 &= \lim_{\theta \rightarrow 1} \omega_5 = \frac{-\beta}{\gamma}, \\
\lim_{\theta \rightarrow 1} \omega_1 &= \lim_{\theta \rightarrow 1} \omega_2 = \lim_{\theta \rightarrow 1} \omega_3 = \lim_{\theta \rightarrow 1} \omega_4 = \frac{\beta}{\alpha}, \\
\lim_{\theta \rightarrow 1} v_0 &= \lim_{\theta \rightarrow 1} v_5 = \frac{(\gamma - \beta)}{\gamma}, \\
\lim_{\theta \rightarrow 1} v_1 &= \lim_{\theta \rightarrow 1} v_2 = \frac{(\beta - \alpha)}{\alpha}, \quad \lim_{\theta \rightarrow 1} v_3 = \lim_{\theta \rightarrow 1} v_4 = \frac{(\beta + \alpha)}{\alpha}, \\
\lim_{\theta \rightarrow 1} r(t) &= \begin{cases} \frac{\alpha \bar{s}}{(\alpha - \beta)}, & \text{if } t_0 \leq t \leq t_1 \\ \bar{s}, & \text{if } t_1 \leq t \leq t_4 \\ \frac{\alpha \bar{s}}{(\alpha + \gamma)}, & \text{if } t_4 < t \leq t_e. \end{cases}
\end{aligned} \tag{26}$$

Substituting (26) into (19), we have

$$\begin{aligned}
t_0 &= -\frac{\gamma}{\beta + \gamma} \cdot \frac{N}{\bar{s}} + \frac{\gamma - \beta}{\beta + \gamma} \delta, \quad t_5 = t_e = \frac{\beta}{\beta + \gamma} \cdot \frac{N}{\bar{s}} + \frac{\gamma - \beta}{\beta + \gamma} \delta, \\
t_1 &= t_2 = -\frac{\beta \gamma}{\alpha(\beta + \gamma)} \cdot \frac{N}{s} + \frac{\beta - \alpha}{\alpha} \delta, \\
t_3 &= t_4 = -\frac{\beta \gamma}{\alpha(\beta + \gamma)} \cdot \frac{N}{s} + \frac{\alpha + \beta}{\alpha} \delta.
\end{aligned} \tag{27}$$

Therefore, we get the same traffic flow pattern with that from a deterministic bottleneck model.  $\square$

**Proposition 5.** *When the number of commuters is given, increasing the value of parameter  $\theta$  will result in a decrease in the length of peak period.*

*Proof.* According to (16), we have  $d\hat{s}/d\theta = \alpha/(\alpha + \gamma) > 0$ . This implies that  $\hat{s}$  is a monotonic increasing function with respect to  $\theta$ . According to (17), we can obtain the length of peak period as follows:

$$t_e - t_0 = \frac{N}{\hat{s}} \frac{\omega_0}{\omega_0 - 1} + \frac{v_0}{1 - \omega_0} \delta - \frac{N}{\hat{s}} \frac{1}{\omega_0 - 1} - \frac{v_0}{1 - \omega_0} \delta = \frac{N}{\hat{s}}. \tag{28}$$

Since  $N$  is constant and  $\hat{s}$  is a monotonic increasing function with respect to  $\theta$ ,  $t_e - t_0$  is also a monotonic increasing function with respect to  $\theta$ . This completes the proof.  $\square$

The above proof also shows that the length of peak period is not affected by the  $\delta$ -value.

### 3. Numerical Examples

The input parameters of our numerical example are  $\alpha = 6.4$  \$/h,  $\beta = 3.9$  \$/h,  $\gamma = 15.21$  \$/h,  $N = 6000$  veh,  $\bar{s} = 4000$  veh/h,  $\delta = 10$  min, and  $\theta = 0.9$ . By solving the proposed

TABLE 1: The influence of parameter  $\theta$  on the mean trip cost and the watershed time instant.

$\theta$	$E[c(t)]$	$t_0$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_e$	$t_e - t_0$
1.00	3.62	-1.10	-0.74	-0.74	-0.40	-0.40	0.40	0.40	1.50
0.95	3.78	-1.14	-0.77	-0.73	-0.46	-0.31	0.36	0.39	1.52
0.90	3.95	-1.18	-0.80	-0.73	-0.52	-0.21	0.31	0.37	1.55
0.85	4.13	-1.23	-0.85	-0.72	-0.59	-0.09	0.26	0.35	1.57
0.80	4.33	-1.28	-0.89	-0.71	-0.66	0.05	0.21	0.32	1.60

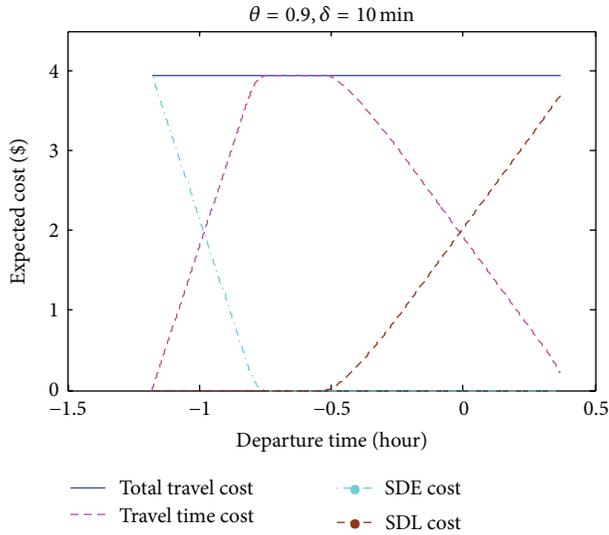


FIGURE 3: The mean equilibrium trip cost and other components.

model, we obtained the time period with positive departure rate during  $[-1.18, 0.37]$  (see Table 1). The mean trip cost, the mean travel time cost, and the mean schedule delay cost (SDE and SDL) are depicted in Figure 3. We find that the mean trip costs of all commuters are the same and are equal to 3.95 \$, but endure a trade-off between cost of travel time and cost of schedule delay. It is interesting that the waiting time cost is not zero at the end of the peak period, indicating that queue still exists.

It is interesting to investigate the impact of the fraction parameter  $\theta$  on the solution of the stochastic bottleneck model. We changed the parameter  $\theta$  from 0.8 to 1.0, and computed the mean trip cost and the watershed time instants. The results are shown in Table 1; the row highlighted is with the default  $\theta$ -value. It can be seen that  $t_1 = t_2, t_3 = t_4$ , and  $t_5 = t_e$  when  $\theta = 1.0$ . This confirms Proposition 4. We can also find that the length of time period with positive departure rate increases with the  $\theta$ -value. This confirms Proposition 5. In addition, the second column of Table 1 shows that the mean trip cost decreases with the increase of  $\theta$ -value. This confirms Theorem 2. Since decreasing the  $\theta$ -value is equivalent to increasing the travel time uncertainty, commuters will leave home earlier than before for avoiding the potential loss caused by uncertainty risk.

The departure rates against different  $\theta$ -values are displayed in Figure 4. It can be clearly seen that the results confirm Propositions 4 and 5 that the stochastic bottleneck model immediately follows the deterministic model when

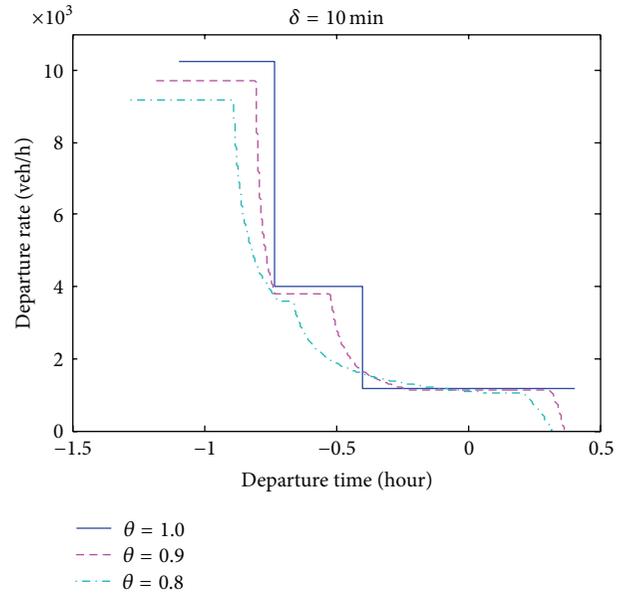


FIGURE 4: The influence of parameter  $\theta$  on departure rate.

$\theta$  approaches one, and that enlarging the parameter  $\theta$  will result in a decrease in the length of peak period. Figure 4 also shows that in equilibrium the departure rate during the peak period is monotonically decreasing, which is consistent with Theorem 3.

Table 2 lists the mean trip costs and the watershed time instants when  $\theta = 0.9$  and  $\delta$ -value from 0 minutes to 20 minutes. It is shown that in equilibrium the mean trip cost is monotonically decreasing with increasing  $\delta$ , whilst the length of the peak period remains unchanged. This is consistent with Theorem 1.

Figure 5 depicts the departure rates against different  $\delta$ -values at  $\theta = 0.9$ . It can be seen that with larger  $\delta$ -value (or longer PATI), the peak periods shift to later and the amount of earlier departures (at a rate higher than capacity) reduces. The overall departure time patterns become more flat and approach designed capacity of the bottleneck with increasing  $\delta$ -value. This suggests that traffic congestion can be alleviated by adopting flexi-time work schedule, similar to that achieved through congestion pricing policy (see, Arnott et al. [5]).

Figure 6 depicts the mean trip times with different preferred arrival time intervals when  $\theta = 0.9$ . One can observe from this figure that adopting flexi-time work schedule can reduce the commuters' travel time or the queue behind the bottleneck. The areas below three curves are 0.6208, 0.5940, and 0.5131 hours, respectively. It shows further that the

TABLE 2: The influence of parameter  $\delta$  on the mean trip cost and the watershed time instants.

$\delta$ min	$E[c(t)]$	$t_0$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_e$	$t_e - t_0$
0	4.98	-1.28	-0.80	-0.80	-0.80	-0.55	0.21	0.27	1.55
5	4.46	-1.23	-0.80	-0.72	-0.68	-0.38	0.26	0.32	1.55
10	3.95	-1.18	-0.80	-0.73	-0.52	-0.21	0.31	0.37	1.55
15	3.43	-1.13	-0.80	-0.74	-0.36	-0.04	0.36	0.42	1.55
20	2.91	-1.08	-0.80	-0.75	-0.20	0.13	0.41	0.47	1.55

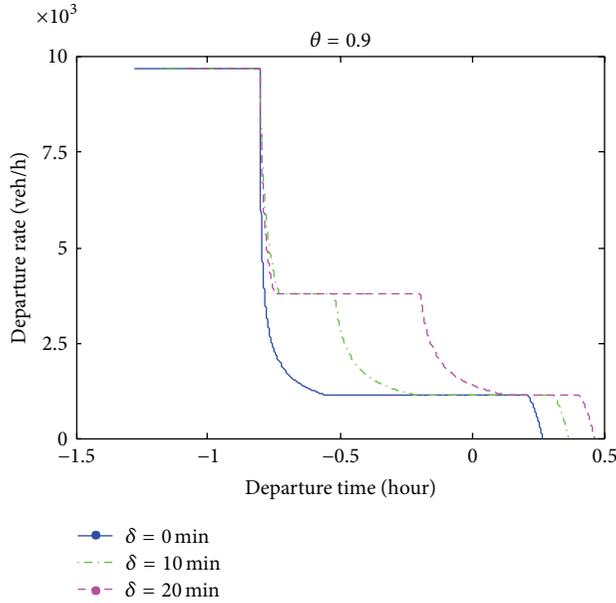


FIGURE 5: The influence of parameter  $\delta$  on departure rate.

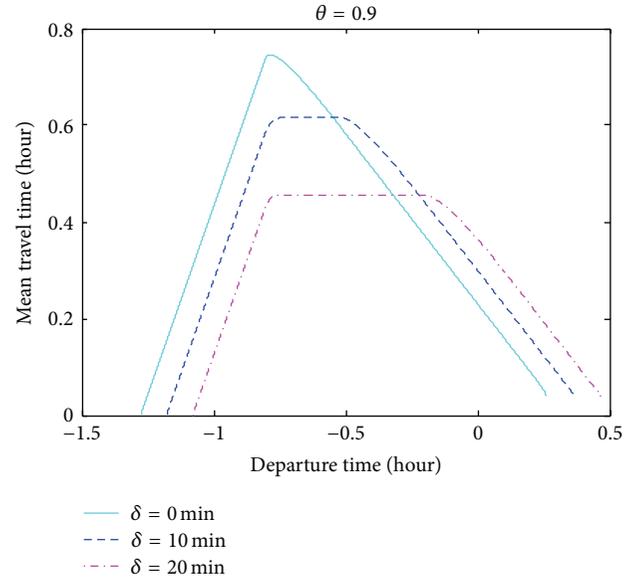


FIGURE 6: Travel time with different  $\delta$ -value.

adoption of flexi-time in equilibrium leads to less congestion than under the fixed work schedule.

Figure 7 shows the joint effect of the  $\theta$ -value and the  $\delta$ -value on the equilibrium trip cost. For a fixed  $\theta$ , the expected trip cost declines with an increase of the preferred arrival time interval. Since the schedule delay costs endured by commuters are reduced with an increase of the length PATI without changing the departure pattern, this means that transport policies to encourage firms in CBD to adopt flexi-time can ease overall system traffic congestion. On the other hand, for a fixed  $\delta$ -value, the expected trip cost declines with the decrease of  $\theta$ -value, which confirms that improving system reliability and reducing uncertainty will increase the system's effectiveness.

### 4. Conclusion

This paper investigated the travel choice behavior under uncertainty on morning commute problem by considering the capacity variability of a highway bottleneck. The bottleneck model was applied to analyze the departure time pattern of a group of homogeneous commuters with the same preferred arrival time interval. The capacity of the bottleneck is assumed to follow a uniform distribution and

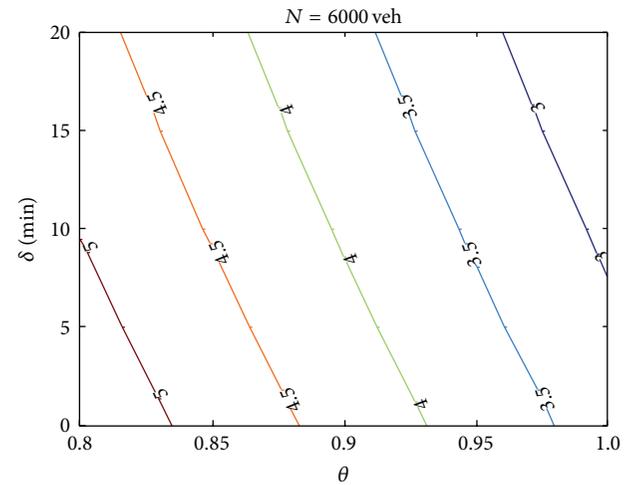


FIGURE 7: The equilibrium trip cost  $E[C]$  with different  $\theta$  and  $\delta$ .

the commuters' departure time choice to follow the UE principle in terms of the mean trip cost. The analytical solution of the stochastic bottleneck model was derived. Both analytical and numerical results show that increasing the capacity variation results in longer peak period and higher commuters' mean trip cost. In addition, it is shown that with longer flexi-time interval, the departure time distributions

become flatter. This suggests that flexi-time is an effective demand management measure for alleviating peak congestion. For future research, we will further improve the model with consideration of heterogeneous commuters and travel risk and apply the model in analyzing such policy measures as congestion pricing, metering, and flexible work scheme.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

The authors acknowledge thank the financial support from the National Basic Research Program of China (2012CB725401), the PhD Student Innovation Fund of Beihang University (302976), and the China Scholarship Council.

### References

- [1] W. S. Vickrey, "Congestion theory and transport investment," *American Economic Review*, vol. 59, pp. 251–261, 1969.
- [2] M. J. Smith, "The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck," *Transportation Science*, vol. 18, no. 4, pp. 385–394, 1984.
- [3] C. F. Daganzo, "Uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck," *Transportation Science*, vol. 19, no. 1, pp. 29–37, 1985.
- [4] R. M. Braid, "Uniform versus peak-load pricing of a bottleneck with elastic demand," *Journal of Urban Economics*, vol. 26, no. 3, pp. 320–327, 1989.
- [5] R. Arnott, A. de Palma, and R. Lindsey, "Economics of a bottleneck," *Journal of Urban Economics*, vol. 27, no. 1, pp. 111–130, 1990.
- [6] H. Yang and H.-J. Huang, "The multi-class, multi-criteria traffic network equilibrium and systems optimum problem," *Transportation Research Part B*, vol. 38, no. 1, pp. 1–15, 2004.
- [7] G. Ramadurai, S. V. Ukkusuri, J. Zhao, and J.-S. Pang, "Linear complementarity formulation for single bottleneck model with heterogeneous commuters," *Transportation Research Part B*, vol. 44, no. 2, pp. 193–214, 2010.
- [8] M. J. Machina, "Choice under uncertainty: problem solved and unsolved," *Journal of Economic Perspectives*, vol. 1, no. 1, pp. 121–154, 1987.
- [9] R. P. Cubitt and R. Sugden, "Dynamic decision-making under uncertainty: an experimental investigation of choice between accumulator gambles," *Journal of Risk and Uncertainty*, vol. 22, no. 2, pp. 103–128, 2001.
- [10] C. Gollier and N. Treich, "Decision-making under scientific uncertainty: the economics of the precautionary principle," *Journal of Risk and Uncertainty*, vol. 27, no. 1, pp. 77–103, 2003.
- [11] M. H. Birnbaum and U. Schmidt, "An experimental investigation of violations of transitivity in choice under uncertainty," *Journal of Risk and Uncertainty*, vol. 37, no. 1, pp. 77–91, 2008.
- [12] C. Starmer, "Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk," *Journal of Economic Literature*, vol. 38, no. 2, pp. 332–382, 2000.
- [13] A. de Palma, M. Ben-Akiva, D. Brownstone et al., "Risk, uncertainty and discrete choice models," *Marketing Letters*, vol. 19, no. 3-4, pp. 269–285, 2008.
- [14] J. V. Henderson, "The economics of staggered work hours," *Journal of Urban Economics*, vol. 9, no. 3, pp. 349–364, 1981.
- [15] S.-I. Mun and M. Yonekawa, "Flexitime, traffic congestion and urban productivity," *Journal of Transport Economics and Policy*, vol. 40, no. 3, pp. 329–358, 2006.
- [16] B. Siu and H. K. Lo, "Equilibrium trip scheduling in congested traffic under uncertainty," in *Proceedings of the 18th International Symposium on Transportation and Traffic Theory*, W. H. K. Lam, S. C. Wong, and H. K. Lo, Eds., pp. 19–38, Elsevier, Oxford, UK, 2009.
- [17] H. Li, M. Bliemer, and P. Bovy, "Departure time distribution in the stochastic bottleneck model," *International Journal of ITS Research*, vol. 6, no. 2, pp. 79–86, 2008.
- [18] L. L. Xiao, H. J. Huang, and R. Liu, "Congestion behavior and tolls in a bottleneck model with stochastic capacity," *Transportation Science*, 2013.
- [19] A. Chen, Z. Ji, and W. Recker, "Travel time reliability with risk-sensitive travelers," *Transportation Research Record*, no. 1783, pp. 27–33, 2002.

## Research Article

# Modeling the Equilibrium Bus Line Choice Behavior and Transit System Design with Oblivious Users

**Chuan-Lin Zhao and Hai-Jun Huang**

*School of Economics and Management, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Chuan-Lin Zhao; chelone@163.com

Received 15 November 2013; Accepted 1 January 2014; Published 6 February 2014

Academic Editor: Huimin Niu

Copyright © 2014 C.-L. Zhao and H.-J. Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In most of transportation literature, users are assumed to be perfectly rational in minimizing their own travel costs or perceived travel costs. However, users may not be perfectly rational in implementing their choices in reality. There exists a kind of boundedly rational users, that is, oblivious users. These oblivious users make their route choices by simple criteria, for example, selecting the shortest (or the most direct) route only based on physical distance or simply following routes recommended by a GPS system. This paper investigates how the existence of oblivious users affects the equilibrium bus line choice behavior in a public transit system. And we propose a method to design a more realistic system.

## 1. Introduction

The purpose of this paper is duple to advance our understanding on the boundedly rational behavior of public transit users when choosing bus lines in a transit network and to design a more realistic public transit system when considering the boundedly rational users.

In the literature, user equilibrium models play an important role in the traffic assignment problems. By assuming that all road users behave in a completely rational way and seeking to minimize their own disutility, Wardrop [1] defined a state of route choice, the so-called user equilibrium (UE). At the UE state, no user can further improve her or his utility by unilaterally changing routes. By relaxing some of the behavioral restrictions implied in a strict deterministic disutility minimization rule, Daganzo and Sheffi [2] developed a stochastic user equilibrium (SUE) model that considers the travelers' imperfect perceptions of travel times. The SUE is achieved when users can no longer change their perceived utility. Existence and uniqueness of UE or SUE in general networks have been well investigated in the literature, including the solution methods for obtaining these two states; see Sheffi [3] and Yang and Huang [4] for more details.

The third equilibrium type is boundedly rational user equilibrium (BRUE). As a relaxation of perfect rationality

and optimality, the notion of bounded rationality was proposed by Simon [5] and introduced to traffic modeling by Mahmassani and Chang [6]. It has been shown that bounded rationality is important in many contexts (see, e.g., Conlisk [7] and references cited therein). In the transportation field, Mahmassani and Chang [6] studied the existence, uniqueness, and stability properties of BRUE in the standard single-link bottleneck network. Many simulation and experimental studies have incorporated the travelers' boundedly rational behavior (e.g., Hu and Mahmassani [8], Mahmassani and Liu [9], and Mahmassani [10]). Lou et al. [11] are the first to systematically examine the mathematical properties of BRUE in a network traffic assignment context. More specifically, as reported in Mahmassani and Chang [6] and discussed by Lou et al. [11] and Di et al. [12], BRUE flow distributions in a static network may not be unique and the set of all possible BRUE flow distributions is a nonconvex and nonempty set.

Recently, Karakostas et al. [13] extended the traditional UE models by considering one kind of boundedly rational users, that is, oblivious users. These users decide their routing only according to the shortest paths observed on a map.

The above studies are only subject to private car systems. In this paper, we will proceed to our study in a public transit system. In the microeconomic analysis of urban public transportation, two types of resources have to be taken into

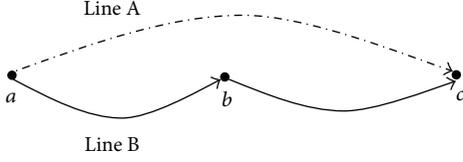


FIGURE 1: A transit network with two lines.

account: those provided by operators, such as vehicles, fuel, terminals, or labor, and those provided by users, namely, their time, usually divided into waiting, access for, and in-vehicle times. In addition, Kraus [14], Lam et al. [15], and Huang et al. [16] introduced the concept of body crowding cost which is related to the passenger/capacity ratio. After Vickrey's view [17], Mohring [18] constructed a microeconomic model to determine optimal frequency of buses serving a corridor with fixed demand. The main result was that frequency should be proportional to the square root of demand, and this happened only because all resources (operators' and users' cost) were considered when finding the minimum cost operation. Jara-Díaz et al. [19] analyzed and compared the total value of the resources consumed (operators' and users' time) under four line structures. The role of users' costs was shown to be crucial. This approach has evolved along the last decades, improving our understanding on public transport operations. These studies are based on a strict hypothesis that all users are perfectly rational. In reality, however, most of the users may be boundedly rational in choosing the bus lines.

There is a kind of users, called oblivious users, who make their bus lane choices without caring about the delay and in-carriage congestion experienced. Their decisions rely on simple criteria, for example, finding the most direct line from the transit map. Recently, Raveau et al. [20], through analyzing actual data, verified such an observation that between two routes with identical trip users prefer the most direct route to indirect one with transfer. They also found that the perceptions of transport users regarding available route alternatives are such that they do not always choose what the modeler would consider as the "lowest cost" option. Inspired by this finding, we are naturally interested in the following questions: how the existence of oblivious users affects the equilibrium bus line choice behavior in public transit system and how to design a more realistic transit system when oblivious users exist.

The remainder of this paper is organized as follows. In Section 2, we model the equilibrium bus line choice behavior in a transit network with oblivious users. In Section 3, considering the design variable, that is, bus frequency, we design the system. Section 4 concludes the paper.

## 2. Problem Formulation

In order to explore the analytical results, we proceed to the study in a simple transit network as shown in Figure 1.

In Figure 1, node  $a$  represents the residential zone generating  $N$  commuters to the central business district (CBD) at node  $c$ . Suppose that at nodes  $a$  and  $b$ , there are buses departing from these two nodes.

Under this setting, two bus lines, Line A and Line B, are designed to serve the demand  $N$ . In Figure 1, the chain dotted curve is Line A and the solid curve is Line B. Line A sends buses from  $a$  to  $c$  directly. Line B has a transfer stop at node  $b$ . The transfer is needed for users who choose Line B.

We assume that the percentage of oblivious users among all commuters is  $k$  ( $0 \leq k \leq 1$ ). Following the empirical study of Raveau et al. [20], we further assume that all oblivious users only choose Line A and other users are rational to choose one of the two lines according to the user equilibrium principle.

The total cost  $C^i$  experienced by a commuter who travels from  $a$  to  $c$  by choosing Line  $i$ ,  $i = \{A, B\}$ , can be formulated as

$$C^i = P_w t_w^i + P_v t_v^i + P_c^i + p^i, \quad (1)$$

where  $P_w$  and  $P_v$  are the prices of waiting and in-vehicle times, respectively,  $t_w^i$  and  $t_v^i$  are the average waiting time at bus stops and in-vehicle time associated with Line  $i$ , respectively,  $P_c^i$  is the body congestion cost occurring in bus carriage for Line  $i$ , and  $p^i$  is the transit fare of Line  $i$ .

The body congestion cost is formulated by

$$P_c^i = \begin{cases} 0, & 0 \leq \tau \leq 1, \\ g(\tau), & \tau \geq 1, \end{cases} \quad (2)$$

where  $\tau$  is the passenger/capacity ratio. For obtaining the analytical results, we assume in this paper that  $g(\tau)$  is a linearly increasing function of the ratio; that is,  $g(\tau) = \theta\tau$ . For  $0 \leq \tau \leq 1$ , all passengers can find seats and the body congestion does not exist, so  $P_c^i = 0$ . In this paper, the capacity is simply defined as the total number of seats provided by a bus line.

For a constant arrival rate of passengers and regular bus headways, the average waiting time  $t_w^A$  for the commuters who choose Line A is

$$t_w^A = \frac{1}{2f^A}, \quad (3)$$

where  $f^A$  is the bus frequency of Line A.

Because the commuters who choose Line B have to transfer at node  $b$ , their average waiting time is

$$t_w^B = \frac{1}{f^B}, \quad (4)$$

where  $f^B$  is the bus frequency of Line B.

The in-vehicle time includes the bus running time, the time waiting for other commuters' boarding at origin, and the time waiting for other commuters' alighting at destination. For different bus lines, the in-vehicle time is different.

For Line A, the in-vehicle time is

$$t_v^A = \frac{T_0^A}{2} + \frac{tN^A}{f^A}, \quad (5)$$

where  $T_0^A$  is the bus cycle running time of Line A,  $t$  is the average boarding and alighting time for each commuter, and  $N^A/f^A$  is the number of passengers boarding each bus of

Line A. In (5), the first term of the right-hand side is the bus moving time and the second term is the boarding and alighting time caused by all passengers sequentially alighting and boarding.

For Line B, there is a transfer at node  $b$  and the in-vehicle time is

$$t_v^B = \frac{T_0^B}{2} + \frac{2tN^B}{f^B}. \quad (6)$$

An equilibrium state is reached when all commuters are satisfied with their bus line choice. In other words,  $kN$  oblivious users choose Line A, and other rational users experience identical and minimal travel cost no matter they choose which line.

For facilitating the presentation of the essential idea, we assume that the passenger/capacity ratio of each line is larger than 1 and  $P_w = P_v = P$ . Based on these assumptions, the perfectly rational user equilibrium solution  $(\bar{N}^A, \bar{N}^B)$ , that is, when  $C^A = C^B$  and  $k = 0$ , is easily found. It follows that

$$\begin{aligned} \bar{N}^A &= (2f^A V^A N (2PtV^B + \theta) - PV^A V^B (f^B - 2f^A) \\ &\quad + (p^B - p^A + PT_0^B - PT_0^A) f^A V^A f^B V^B) \\ &\quad \times (2f^B V^B (PtV^A + \theta) + 2f^A V^A (2PtV^B + \theta))^{-1}, \\ \bar{N}^B &= N - \bar{N}^A, \end{aligned} \quad (7)$$

where  $V^i$  is the bus capacity of Line  $i$ ,  $i = \{A, B\}$ .

Clearly, the above bus line split solution is affected by the bus cycle running time, bus frequency, capacity, and fare. Next, we derive some results in three special cases.

*Case I* ( $V^A = V^B = V$ ,  $p^A = p^B = p$ , and  $f^A \neq f^B$ ). Equation (7) becomes

$$\begin{aligned} \bar{N}_1^A &= (2f^A N (2PtV + \theta) - PV (f^B - 2f^A) \\ &\quad + (PT_0^B - PT_0^A) f^A f^B V) \\ &\quad \times (2f^B (PtV + \theta) + 2f^A (2PtV + \theta))^{-1}, \\ \bar{N}_1^B &= N - \bar{N}_1^A. \end{aligned} \quad (8)$$

*Case II* ( $f^A = f^B = f$ ,  $p^A = p^B = p$ , and  $V^A \neq V^B$ ). Equation (7) becomes

$$\begin{aligned} \bar{N}_2^A &= \frac{2V^A N (2PtV^B + \theta) + P(1 + T_0^B - T_0^A) f V^A V^B}{2V^B (PtV^A + \theta) + 2V^A (2PtV^B + \theta)}, \\ \bar{N}_2^B &= N - \bar{N}_2^A. \end{aligned} \quad (9)$$

*Case III* ( $f^A = f^B = f$ ,  $V^A = V^B = V$ , and  $p^A \neq p^B$ ). Equation (7) becomes

$$\begin{aligned} \bar{N}_3^A &= \frac{2N (2PtV + \theta) + PV + (p^B - p^A + PT_0^B - PT_0^A) f V}{2(PtV + \theta) + 2(2PtV + \theta)} \\ \bar{N}_3^B &= N - \bar{N}_3^A. \end{aligned} \quad (10)$$

When oblivious users are considered, that is,  $k > 0$ , the following proposition can be obtained.

**Proposition 1.** *If  $kN \leq \bar{N}^A$ , the oblivious users do not affect the equilibrium choice state; otherwise, the equilibrium choice state becomes  $(kN, (1 - k)N)$ .*

*Proof.* Substituting (2) to (6) into (1) yields

$$\begin{aligned} C^A &= P \frac{1}{2f^A} + P \left( \frac{T_0^A}{2} + \frac{tN^A}{f^A} \right) + \theta \frac{N^A}{f^A V^A} + p^A, \\ C^B &= P \frac{1}{f^B} + P \left( \frac{T_0^B}{2} + \frac{2tN^B}{f^B} \right) + \theta \frac{N^B}{f^B V^B} + p^B. \end{aligned} \quad (11)$$

Obviously, in (11) the total cost experienced by a commuter choosing Line A (or B) is linearly increasing with respect to the number of passengers choosing Line A (or B).

Consider a corner of the initial state that all oblivious users choose Line A and all rest rational users choose Line B.

If  $kN \leq \bar{N}^A$ , some rational users who have higher cost in Line B will switch to Line A until the equilibrium state  $(\bar{N}^A, \bar{N}^B)$  is achieved.

If  $kN > \bar{N}^A$ , the oblivious users who choose Line A are satisfied with their choice; meanwhile the rest rational users who choose Line B incur lower cost and will not change their choice. So the equilibrium state is  $(kN, (1 - k)N)$ .  $\square$

In this proposition, the parameter  $k$  is crucial and can be calibrated by empirical survey (e.g., stated preference survey).

### 3. Transit System Design

In this section, we only consider the situation that  $kN > \bar{N}^A$  in which the equilibrium state  $(kN, (1 - k)N)$  is affected by oblivious users. In order to find the design variable, for example, frequency  $f^i$ , the total value of the resources (VRC) consumed by operators and users per hour has to be minimized as shown in (12). The first term on the right-hand side of (12) corresponds to the operation cost of fleet sizes times operational costs. The fleet of each line results from the product of frequency  $f^i$  times its cycle time  $t_c^i$ . In the operational cost we consider a constant term  $c_0$  and a variable term that grows with the size of the vehicle,  $c_1 V^i$ . The second term on the right-hand side of (12) represents the total users' costs including the waiting costs, in-vehicle costs, and body congestion cost. Note that the total cost should not include the transit fare. We then have a minimization problem for each line as follows:

$$\min \text{VRC}^i = f^i t_c^i (c_0 + c_1 V^i) + (P_w t_w^i + P_v t_v^i + P_c^i) N^i, \quad (12)$$

$i = \{A, B\}$ ,

where  $N^A = kN$  and  $N^B = (1 - k)N$ . The cycle time  $t_c^i$  is the summation of vehicle running time and standing time at stops. For Line A, the running time is  $T_0^A$ . The standing time is given by the delay caused by sequentially alighting and boarding time of a passenger ( $t$ ) times the number of passengers boarding each bus ( $N^A/f^A$ ). The cycle time for Line A can be written as

$$t_c^A = T_0^A + \frac{tN^A}{f^A}. \quad (13)$$

Similarly, the cycle time for Line B is

$$t_c^B = T_0^B + \frac{2tN^B}{f^B}. \quad (14)$$

Solving the minimization problem (12), we have the following proposition.

**Proposition 2.** For Lines A and B, suppose that the parameters  $c_0$ ,  $c_1$ ,  $T_0^A$ ,  $T_0^B$ ,  $\theta$ , and  $t$  are constant and the vehicle size is determined by the demand; that is,  $V = N/f$ . Then, the optimal bus frequencies for each line, respectively, are

$$\bar{f}^A = \sqrt{\frac{2t(c_1 + P_v)(kN)^2 + P_w kN}{2T_0^A c_0}}, \quad (15)$$

$$\bar{f}^B = \sqrt{\frac{2t(c_1 + P_v)((1 - k)N)^2 + P_w(1 - k)N}{T_0^B c_0}}. \quad (16)$$

*Proof.* For Line A, the VRC can be written as a function of frequency only as follows:

$$\text{VRC}^A = T_0^A c_0 f^A + \frac{1}{f^A} \left( 0.5P_w N^A + t(P_v + c_1)(N^A)^2 \right) + M, \quad (17)$$

where  $M$  is a constant term,  $M = (T_0^A c_1 + t c_0 + 0.5T_0^A P_v + \theta)N^A$ , and  $N^A = kN$ .

Equation (17) shows that increasing frequency has a double effect. Taking a derivative of (17) with respect to frequency, we have

$$\frac{d\text{VRC}^A}{df^A} = T_0^A c_0 - \frac{1}{(f^A)^2} \left( 0.5P_w N^A + t(P_v + c_1)(N^A)^2 \right). \quad (18)$$

Making it equal to zero and noting that the second derivative is positive, we get the optimal frequency of Line A:

$$\bar{f}^A = \sqrt{\frac{2t(c_1 + P_v)(kN)^2 + P_w kN}{2T_0^A c_0}}. \quad (19)$$

Similarly, we can get the optimal bus frequency of Line B, as given by (16).  $\square$

## 4. Conclusions

In this paper, we studied the equilibrium bus line choice behavior with oblivious users and investigated how the equilibrium state is affected by these users. We further optimized each line's bus frequency of the transit system. Oblivious users are those who stubbornly adhere to some options, regardless of actual conditions. Obviously, such users or passengers indeed exist in reality. Hence, we have to consider them when formulating the option choice model. Our on-going work is to calibrate the model parameters and extend the proposed approach to explore the types of behavior in more complex transit networks.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

The work described in this paper was supported by a Grant from the National Basic Research Program of China (2012CB725401).

## References

- [1] J. Wardrop, "Some theoretical aspects of road traffic research," in *Proceedings of the Institution of Civil Engineers Part II*, pp. 325–378, 1952.
- [2] C. F. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transportation Science*, vol. 11, no. 3, pp. 253–274, 1977.
- [3] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.
- [4] H. Yang and H. J. Huang, *Mathematical and Economic Theory of Road Pricing*, Elsevier, Oxford, UK, 2005.
- [5] H. A. Simon, "A behavioral model of rational choice," *The Quarterly Journal of Economics*, vol. 69, no. 1, pp. 99–118, 1955.
- [6] H. S. Mahmassani and G.-L. Chang, "On boundedly rational user equilibrium in transportation systems," *Transportation Science*, vol. 21, no. 2, pp. 89–99, 1987.
- [7] J. Conlisk, "Why bounded rationality?" *Journal of Economic Literature*, vol. 34, no. 2, pp. 669–700, 1996.
- [8] T.-Y. Hu and H. S. Mahmassani, "Day-to-day evolution of network flows under real-time information and reactive signal control," *Transportation Research C*, vol. 5, no. 1, pp. 51–68, 1997.
- [9] H. S. Mahmassani and Y.-H. Liu, "Dynamics of commuting decision behaviour under Advanced Traveller Information Systems," *Transportation Research C*, vol. 7, no. 2-3, pp. 91–107, 1999.
- [10] H. S. Mahmassani, "Trip timing," in *Handbook of Transport Modeling*, D. A. Hensher and K. J. Button, Eds., Elsevier, New York, NY, USA, 2000.
- [11] Y. Lou, Y. Yin, and S. Lawphongpanich, "Robust congestion pricing under boundedly rational user equilibrium," *Transportation Research B*, vol. 44, no. 1, pp. 15–28, 2010.
- [12] X. Di, H. Liu, J. Pang, and X. Ban, "Boundedly rational user equilibria (BRUE): mathematical formulation and solution sets," *Transportation Research B*, vol. 57, pp. 300–313, 2013.

- [13] G. Karakostas, T. Kim, A. Viglas, and H. Xia, "On the degradation of performance for traffic networks with oblivious users," *Transportation Research B*, vol. 45, no. 2, pp. 364–371, 2011.
- [14] M. Kraus, "Discomfort externalities and marginal cost transit fares," *Journal of Urban Economics*, vol. 29, no. 2, pp. 249–259, 1991.
- [15] W. H. K. Lam, C.-Y. Cheung, and C. F. Lam, "A study of crowding effects at the Hong Kong light rail transit stations," *Transportation Research A*, vol. 33, no. 5, pp. 401–415, 1999.
- [16] H. J. Huang, Q. Tian, and Z. Y. Gao, "An equilibrium model in urban transit riding and fare policies," in *Algorithmic Applications in Management*, vol. 3521 of *Lecture Notes in Computer Science*, pp. 112–121, 2005.
- [17] W. Vickrey, "Some implications of marginal cost pricing for public utilities," *American Economic Review*, vol. 45, no. 2, pp. 605–620, 1955.
- [18] H. Mohring, "Optimization and scale economies in urban bus transportation," *American Economic Review*, vol. 62, no. 4, pp. 591–604, 1972.
- [19] S. R. Jara-Díaz, A. Gschwender, and M. Ortega, "Is public transport based on transfers optimal? A theoretical investigation," *Transportation Research B*, vol. 46, no. 7, pp. 808–816, 2012.
- [20] S. Raveau, J. C. Muñoz, and L. de Grange, "A topological route choice model for metro," *Transportation Research A*, vol. 45, no. 2, pp. 138–147, 2011.

## Research Article

# Statistical Analysis of Traffic Accidents in Shanghai River Crossing Tunnels and Safety Countermeasures

Linjun Lu,<sup>1</sup> Jian Lu,<sup>2</sup> Yingying Xing,<sup>1</sup> Chen Wang,<sup>1</sup> and Fuquan Pan<sup>3</sup>

<sup>1</sup> School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup> Department of Civil and Environmental Engineering, University of South Florida, Tampa, FL 33620, USA

<sup>3</sup> School of Automobile and Transportation, Qingdao Technological University, Qingdao, Shandong 266520, China

Correspondence should be addressed to Jian Lu; jlu2@usf.edu

Received 14 September 2013; Revised 21 November 2013; Accepted 4 December 2013; Published 4 February 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Linjun Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A large number of traffic tunnel accidents have been reported in China since the 21st century. However, few studies have been reported to analyze traffic accidents that have occurred in urban road tunnels. This study aims to examine the characteristics of the temporal, spatial, and modality distributions of traffic in Shanghai river crossing tunnels using statistical analysis and comparative analysis. Employing these techniques tunnel accident data obtained from Shanghai center I10 was analyzed to determine temporal and spatial distribution characteristics of traffic accidents in river crossing tunnels in Shanghai. The results of this analysis are discussed and summarized in this paper. Identification of the characteristics of tunnel traffic accidents can provide valuable information for development of effective countermeasures to improve tunnel safety in China.

## 1. Introduction

Tunnels play an important role in urban transportation networks in large cities in China. As shown in Figure 1 [1], the introduction of new of road tunnels in China has rapidly increased since the beginning of the 21st century. According to a 2012 Chinese government bulletin, by end of 2011 there were 8522 road tunnels in China with total length of about 6.2534 million meters. In 2011, 1138 road tunnels with total length of 1.1309 million meters were constructed. As the number of tunnels has increased, so has the number of traffic accidents in these tunnels. In comparison to open roadways, traffic accidents in long tunnels are relatively rare, but the complex problems of a tunnel accident result in casualties, property damage, and greater social impact [2–5]. In comparison to open roadways, tunnels have some unique characteristics due to their confined nature. Consequently, when a traffic accident occurs, a tunnel can quickly become a fatal trap for motorists due to poisonous smoke inhalation and high temperatures if the accident results in a fire [6, 7].

A number of research studies have been conducted to study traffic tunnel safety. One such study, conducted by

Amundsen and Ranes in 2000 [8] considered traffic accidents in 900 road tunnels on national and county roads. These author's results showed that tunnels had similar safety performance when compared with high standard modern roads. However, at tunnel the entrance zone accident rates were higher and more severe when compared to surface roads. Similarly, the average frequency of fires in tunnels was found to be higher than on open roads. In addition, in some tunnels the frequency of fires involving heavy commercial vehicles was much higher than that of passenger cars according to the research in 1998 by PIARC [9]. Although accident rates appeared to be slightly lower in tunnels than on open roads, accidents that occur in tunnels could have greater peripheral impact [10]. In contrast to the prevailing opinion, a recent study concluded that accident rates were lower in tunnels than on open roads and accidents with severe injuries and property damages in tunnels were less frequent than those on open roads [11]. Haack [12] agreed with these conclusions claiming that current safety issues in tunnels can lead to improved vehicle control states resulting in safety improvements on tunnel roads.

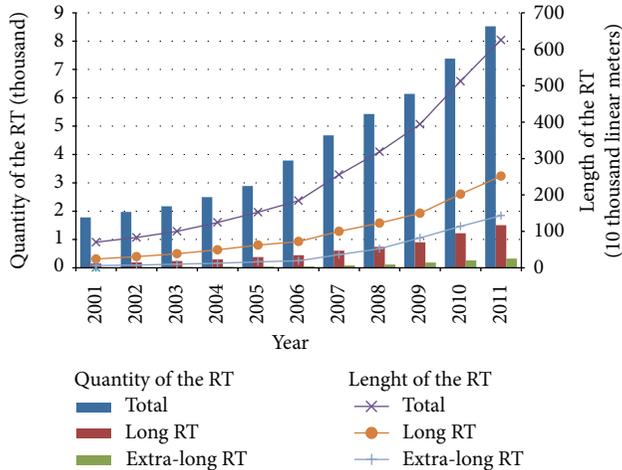


FIGURE 1: Recent road tunnel (RT) development in China 2001~2011.

In China, with the rapid development of tunnel construction and subsequent high rate of tunnel accidents, tunnel traffic safety has acquired greater attention from traffic safety engineers. In 2006, traffic accidents in the Shaoguan tunnel in Jingzhu were examined and safety countermeasures were proposed by Zhang et al. [13]. In 2008, based on a statistical analysis of traffic accidents in four tunnels in Beijing-Zhuhai freeway, temporal, spatial, and modality distribution and accident vehicle type distribution were discussed by Ma et al. [14]. In 2009, Lin et al. [15] used tunnel traffic detector data to examine the spatial and temporal variations of capacity, free-flow speed, passenger car equivalent of buses, and speed-flow relationships. In 2009, Zhang et al. [16] discussed the relationships between traffic accidents, fire accidents, rear-end accidents, wall colliding accidents, and vehicle roll-over accidents with traffic volumes and some countermeasures were suggested to decrease traffic accidents in expressway tunnels. In 2010, Ma et al. [17] conducted logistic model research to determine the factors which affect accident severity in highway tunnels. It was found that the time of the accident, the collision type, the weather conditions, and the ratio of daily PCU to AADT are the most significant factors in the severity of a vehicular accident.

However, despite of these research findings, few studies have been reported to focus on traffic accidents in urban river crossing tunnels. Therefore, using the Shanghai river crossing tunnels as the subject we employed statistical and contrastive analyses to define the temporal distribution, spatial distribution and accident type distribution analyzes of accident severity and the relationship between road classifications in Shanghai river-crossing tunnels.

## 2. Data Basis

This study was based on data collected by Shanghai Center 110, which provided details on traffic accidents in traffic tunnels in Shanghai. Each accident record includes detailed information including time, geographic coordinates, location, case category, and a detailed case description. Accidents

in thirteen river crossing tunnels were selected for this study. The supporting data on river crossing tunnels such as tunnel length, lanes, design speed, and road classification were collected from Shanghai Tunnel Engineering & Rail Transit Design and Research Institute.

Using the methods of Amundsen and Ranes, data on accidents and locations were retrieved from the Shanghai Center 110 and were grouped into the following four categories as shown in Figure 2 [8].

Zone 1: the first 50 m in front of the tunnel openings.

Zone 2: the first 50 m inside the tunnel.

Zone 3: the next 100 m inside the tunnel.

Zone 4: the mid-zone, that is, the remainder of the tunnel.

Tunnels shorter than 100 m only have zones 1 and 2; tunnels shorter than 300 m do not have a mid-zone (i.e., zone 4). In this paper, all tunnels are no less than 1000 m and, therefore, have four zones.

## 3. Temporal Distribution of Traffic Accidents

Temporal characteristics of accidents refer to the characteristics of accidents that vary over time. Analyzing the temporal distribution can reveal the trend of accidents but can also give the basis for further research on the causation of those accidents [18]. In addition to annual surveys, temporal characteristics of accidents can also be studied by analyzing the monthly and hourly distribution of traffic accidents. Consequently, this paper will make a detailed study of the temporal characteristics of traffic accidents by analyzing the weekly and hourly distributions of these incidents.

*3.1. Weekly Distribution of Traffic Accidents.* Patterns of city life and daily commuting impact the weekly distribution of tunnel accidents. As is shown in Figure 3, accidents occur more frequently on Monday and Friday while the occurrence of accidents on weekends was lower than on week days. This could be attributed to the pattern of city life and daily commuting in Shanghai. Similar results were also found by researchers in Victoria, Australia, [19] where accidents were more likely to occur on weekdays than weekends.

*3.2. Hourly Distribution of Traffic Accidents.* The hourly profiles of vehicle trips in Shanghai are shown in Figure 4. As can be seen in Figure 4 [20], trip time is concentrated on two peak periods, 7:00~9:00 and 16:00~18:00. The hourly distribution of traffic accidents in Shanghai river tunnels has similar pattern with that of trip time, reflecting daily commuting routines. As can be seen in Figure 5, traffic accidents occur primarily during the daytime and exhibit morning and evening peak times, namely, 7:00~9:00 and 17:00~19:00, which is consistent with trip peak time. Generally speaking, traffic volume during peak hours is significantly higher than at other time periods. To our knowledge, during peak hours, many Chinese travelers tend to adopt high-risk driving behavior in order to reach their destinations on time. Thus,

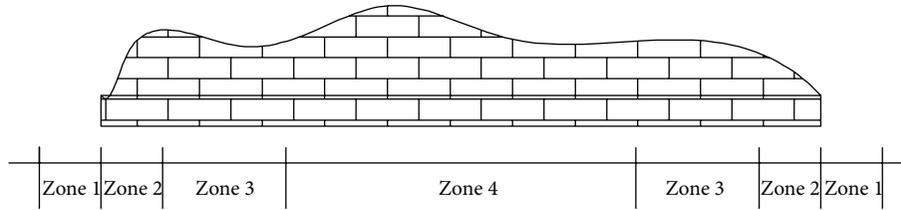


FIGURE 2: Tunnel zones.

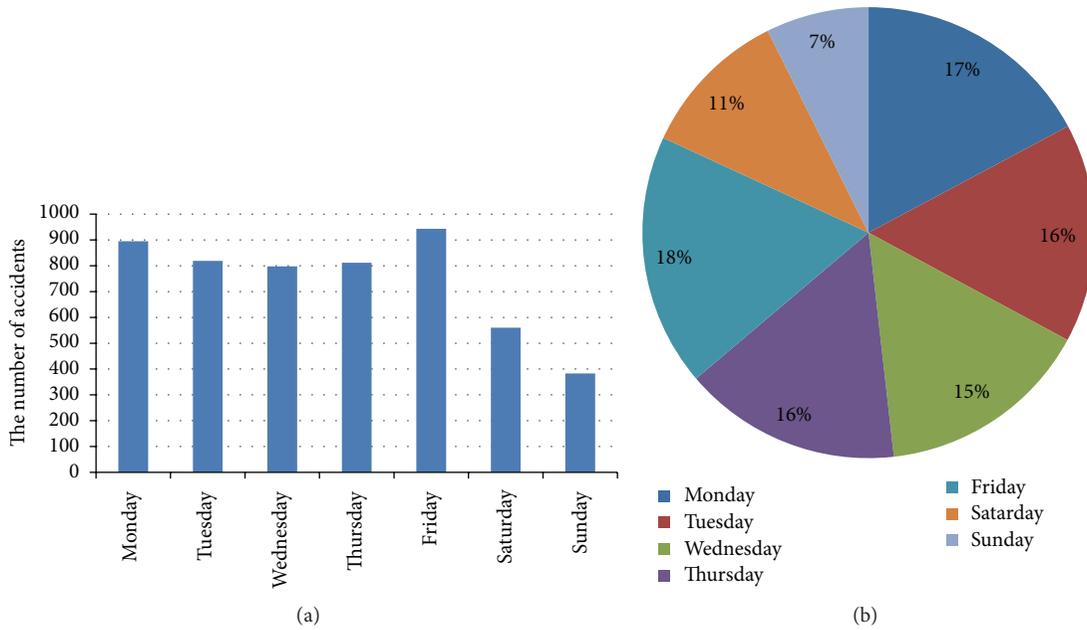


FIGURE 3: Weekly distributions of traffic accidents.

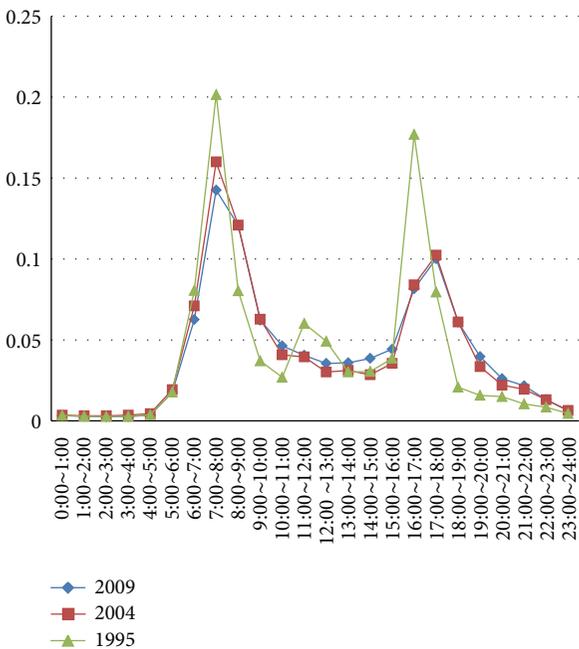


FIGURE 4: Hourly profiles of trips.

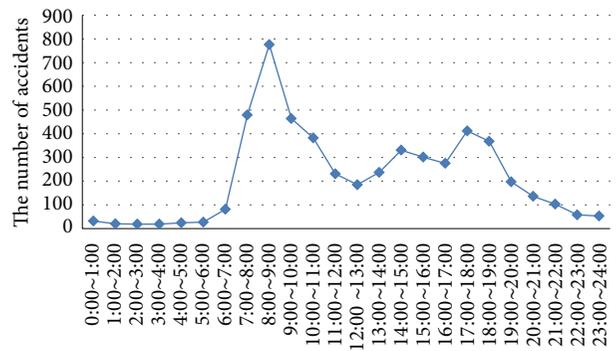


FIGURE 5: Hourly distributions of traffic accidents.

it is very important to enhance traffic management and dispersions during peak hours.

#### 4. Spatial Distributions of Traffic Accidents

4.1. Accident Frequency. Data on accidents and locations retrieved from the Shanghai Center, I10, were grouped into

TABLE 1: Number of injured people per accident.

	Zone 1	Zone 2	Zone 3	Zone 4	Traffic accidents on the city road outside the tunnel	Traffic accidents in tunnel
Number	1.13	1.22	1.36	1.25	1.13	1.26

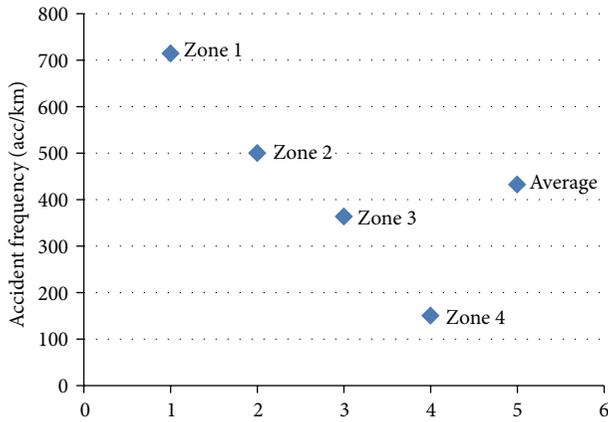


FIGURE 6: Spatial distributions of traffic accidents.

the following four categories. From the summary of tunnel traffic accidents, (Figure 6), the accident frequencies of zone 1 and zone 2 are considerably higher than the average. The accident frequency of zone 3 is slightly below the average and that of zone 4 is significantly lower than the average. Thus, it can be seen that accidents occurred primarily in zones 1, 2, and 3 as opposed to being uniform or randomly distributed. The main reasons of this phenomenon are visibility and speed. Generally speaking, when a vehicle enters a tunnel, the driver requires a short time period to adapt to the dim light conditions (also called “black hole”) in the tunnel. Similarly, when the vehicle leaves the tunnel, intense daylight outside the tunnel will lead to a bright “white hole” both of which adversely affect the drivers vision. In addition, according to Zhao et al.’s [21] research, when a vehicle approaches a tunnel, the driver normally decelerates as he approaches tunnel entrance. After entering the tunnel, the driver will then accelerate to a speed that is lower than that of an open road. These large speed fluctuations have a deleterious impact on traffic safety.

**4.2. Accident Severity Analysis.** A total of 203 persons were injured in the 167 injury-related accidents included in this study. The severity of the injuries in each tunnel zone is shown in Figure 7. Most injury-related traffic accidents occur in zone 1 and zone 4 but there were no fatalities in these accidents. In tunnels the share of serious injuries and fatal injuries in all accidents is 2.4%, while this proportion on the open road outside the tunnel is only 1.2%. A similar pattern is also found for the number of injuries per accident, which is 1.26 in tunnels compared with 1.13 on the city roads outside tunnels. Table 1 gives an overview of the number of injuries per accident.

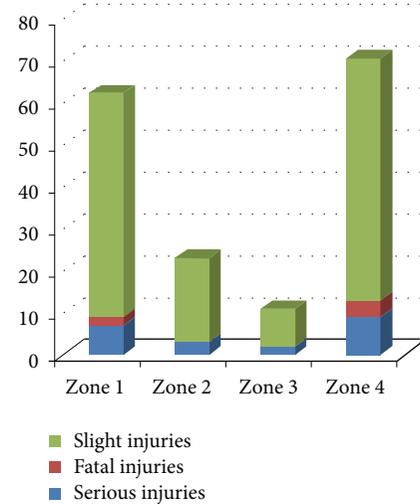


FIGURE 7: Number injured by severity and tunnel zone.

**4.3. Accidents versus Weather.** We also studied the relationship between weather conditions and accidents. Accidents occurring on rainy days stood at 47.6 percent of total number of accidents. Of accidents occurring within 50 m of the opening sections of tunnels (zones 1 and 2), 38.5% took place on rainy days. In zones 3 and 4 the proportions of accidents happening on rainy days was 23.6% and 27%, respectively.

Presumably, the reason for these results is that road friction near tunnel entrances and exits decreases significantly in wet conditions. On rainy days, vehicles drag significant quantities of rainwater into a tunnel which probably reduces the frictional coefficient of the pavement at the entrance zone, thereby leading to an increase in accident risk. When vehicles leave the tunnel, they travel from a dry surface in the tunnel to a slippery road surface near the tunnel exit. This significant change of road friction could also lead to a higher accident risk.

## 5. Accident Types

There is a variety of car accidents, including head-on collisions, sideswipes, rear-end collisions, bumping, scrapes, rollovers, and fires. According to the acquired accident data, accidents in tunnels were divided into seven categories: head-on, sideswipe, rear-end, fixed object, rollover, and others. Figure 8 indicates that rear-end accidents account for 80 percent of all accidents while sideswipe accidents account for only 9 percent. We assume that the basis for these data is that the tunnel roads in Shanghai are all one-way roads with high traffic volumes and most accidents involving multiple vehicles are rear-end or side-swipe accidents. In addition,

TABLE 2: City road classifications and traffic accidents.

Road classification	Number of accidents	Total length of tunnel (km)	Accident frequency (acc/km <sup>-1</sup> )	Design speed (km/h)
Express way	2739	11.023	248.48	80
Arterial road	1162	5.561	208.96	40
Secondary road	1061	17.232	61.57	40

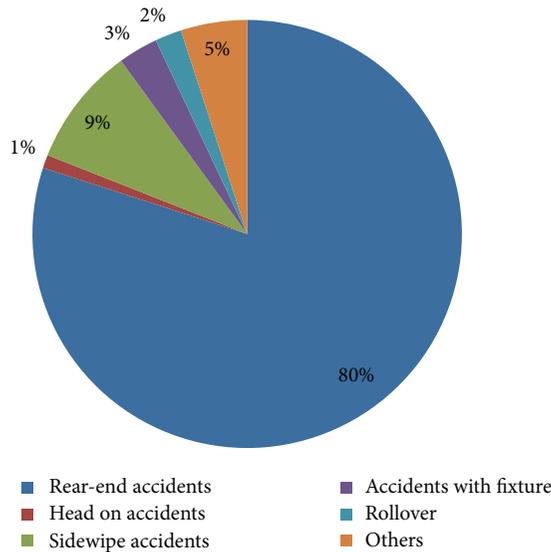


FIGURE 8: Accident types versus traffic accidents.

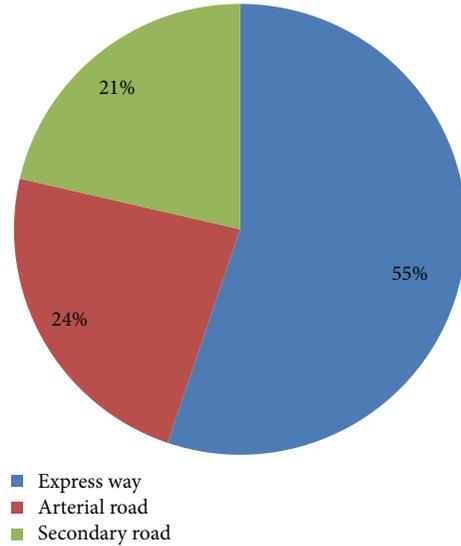


FIGURE 9: Road classifications and traffic accidents.

according to police investigations, tunnel crashes are often caused by drivers' aggressive lane changes and high speed, which leads to rear-end accidents.

### 6. Road Classifications and Traffic Accidents

According to the Code for Transport Planning on Urban Roads [20], urban roads can be divided into the following categories in terms of road functionality: expressway, arterial roads, secondary roads, and slip roads. Traffic volumes and vehicle speed limits vary with the design of the road functionality. Table 2 and Figure 9 give an overview of the relationship between city road classifications and traffic accidents. As can be seen from the figure, most traffic accidents occur on the expressway, accounting for 55 percent of the total, while arterial road and secondary roads account for 24 percent and 21 percent, respectively. At least two fundamental factors appear to contribute to this phenomenon. First, expressway and arterial roads are the traffic arteries of city's transportation network and, therefore, accommodate a very large volume of traffic. This trend will continue as the proliferation of expressways in Shanghai increases currently stretching to over 192 km in length. These expressways constitute 45 percent of road network turnover value. Second, although expressways incorporate access control to prevent pedestrian and nonmotor vehicles traffic, accident rates in expressway tunnels are quite high. The reason for this could be due to the high design speed and operating speed on expressways. As previously stated, transition zones in tunnels are very

dangerous areas. Thus, drivers who are approaching tunnels at high speed are exposed to much higher accident risk.

### 7. Countermeasures

Based upon the described data analysis, suggestions for improving traffic safety in tunnels in Shanghai can be addressed as follows.

- (1) Improvement of road surfaces could be an effective way to reduce traffic accidents in tunnels. First, tunnels should be cleaned regularly to ensure that the surface of the road in the tunnel is free from debris, oil, or any foreign matter. Second, proper drainage design and strict quality control of construction should be used to reduce rainwater runoff into tunnels. In addition, skid-resistance pavement could be installed to improve road friction.
- (2) Traffic control coordinated with traffic guidance can be used as a fundamental method for minimizing tunnel traffic jam problems and help to reduce tunnel traffic accidents. First, installation of variable message signs (VMS) can provide drivers with real-time traffic information so that drivers could change their travel plans when congestions/accidents occur in tunnels. The information on accidents could also warn drivers entering a tunnel to help them avoid secondary accidents. In addition, traffic regulations should be adopted to prohibit high speed vehicle

passing in tunnels. Second, risky driving, aggressive lane change, and speeding are known to be some of the main reasons leading to high accident rates in tunnels. Consequently, establishing reasonable speed limit standards could provide a simple method for controlling risky driving behavior to reduce accidents in tunnels. Third, the presence of large commercial vehicles in tunnels has been shown to cause erratic automobile driving which is directly correlated with traffic accidents. Consequently, traffic regulations should be considered to separate small passenger cars and large vehicles. For example, providing restrictions on large vehicle movement move during certain time periods or restricting them to traveling in designated lanes.

- (3) Horizontal alignment should meet the requirements of design consistency. With regard to geometric design of tunnel sections, curves and ramps need to be avoided as much as possible, especially over long and steep downgrade sections. If ramps and curves cannot be avoided, speed bumps should be installed at specific distances away from tunnel entrances to reduce the speed of tunnel entrance.
- (4) Lighting conditions in zone 1 and zone 2 need to be carefully designed in order to eliminate the “black hole” condition in tunnel entrances and “white hole” conditions at tunnel exits. This will help to alleviate the number of accidents in zone 1 and zone 2 which are considerably higher than in other areas. The tunnel lighting system that is recommended should include entrance section lighting, transition section lighting, the basic lighting of middle section, exit section lighting, and emergency lighting. For the entrance zone of a tunnel, the recommended lighting should start with yellow lights and gradually change to daylight lamps. Such a layout of illumination may help drivers to adapt to the tunnel environment and reduce the effects of “black hole” and “white hole.”

## 8. Conclusion

Patterns of city life and daily commuting exert a major impact on the distribution of traffic accidents in roadway tunnels. Data have shown that the highest incidents of traffic accidents occur on Monday and Friday and the number of traffic accidents on weekends is much lower than on weekdays. The hourly distribution of traffic accidents in tunnels reveals that traffic accidents occur mainly during the daytime and are most significantly during morning and evening peak times, namely, 7:30~9:30 and 16:30~18:30.

Most traffic accidents occur in zone 1, zone 2, and zone 3 in tunnels. The number of traffic accidents near entrances and exits (at zones 1 and 4) are significantly higher than in other tunnel zones. Most injury-related accidents occur in zone 1 and zone 4 but did not result in fatalities. The share of accidents that occurred on rainy day was found to be as much as 47.6% of the overall total number of accidents.

Rear end collisions represent the major type of accident in tunnels. Of the total number of accidents, rear end collisions accounted for the highest percentage among all the accident types, reaching 80%. This is probably due to the failure to maintain a safe distance between vehicles in the tunnel.

The majority of traffic accidents in tunnels resulted from high speed travel (80 km/h or more), possibly due to high speed, heavy traffic volume, and the confined space of the tunnel. It should be noted that the analysis in this paper resulted from a search for the accident patterns rather than a prediction of accident frequency. The analysis of tunnel accidents and related factors may provide a better understanding of tunnel accident risks and the information needed to establish effective safety countermeasures. The safety countermeasures suggested in this paper can be used as a reference for traffic safety improvements for traffic tunnels in Shanghai.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors would like to express special thanks to Shanghai Transport or Port Research Center and Shanghai Center, 110, for providing valuable information on Shanghai River-crossing Tunnel for academic and research activities. This research was supported in part by the National Natural Science Foundation of China (no. 51078232/E0807).

## References

- [1] Comprehensive Planning Department of the Ministry of Transport of People's Republic of China, “2011 Statistics Bulletin about the Development of Domestic Highway & Waterway Transportation Industry of China,” April 2012, [http://www.moc.gov.cn/zhuzhan/tongjigongbao/fenxigongbao/hangyegongbao/201204/t20120425\\_1231778.html](http://www.moc.gov.cn/zhuzhan/tongjigongbao/fenxigongbao/hangyegongbao/201204/t20120425_1231778.html).
- [2] U. Barth, “Managerial and technical aspects of tunnel safety regarding normal and emergency mode,” in *Proceedings of the International Conference Tunnel Safety and Ventilation*, vol. 8, pp. 8–10, Graz, Austria, 2002.
- [3] H. Mashimo, “State of the road tunnel safety technology in Japan,” *Tunnelling and Underground Space Technology*, vol. 17, no. 2, pp. 145–152, 2002.
- [4] NPRA, *Studies on Norwegian Road Tunnels, an Analysis on Traffic Accidents and Car Fires in Road Tunnels*, Norwegian Public Roads Administration, Directorate of Public Roads, 1997.
- [5] UN, “Recommendations of the group of experts on safety in road tunnels—final report,” Document TRANS/AC.7/9, Inland Transport Committee, United Nations Economic and Social Council, 2001.
- [6] J. Modic, “Fire simulation in road tunnels,” *Tunnelling and Underground Space Technology*, vol. 18, no. 5, pp. 525–530, 2003.
- [7] F. Vuilleumier, A. Weatherill, and B. Crausaz, “Safety aspects of railway and road tunnel: example of the Lötschberg railway tunnel and Mont-Blanc road tunnel,” *Tunnelling and Underground Space Technology*, vol. 17, no. 2, pp. 153–158, 2002.

- [8] F. H. Amundsen and G. Ranes, "Studies on traffic accidents in Norwegian road tunnels," *Tunnelling and Underground Space Technology*, vol. 15, no. 1, pp. 3–11, 2000.
- [9] D. Lacroix, "The new PIARC report on fire and smoke control in road tunnels," in *Proceedings of the International Conference on Safety in Road and Rail Tunnels*, pp. 185–197, 1998.
- [10] K. Lemke, "Road safety in tunnels," *Transportation Research Record*, no. 1740, pp. 170–174, 2000.
- [11] A. N. Beard and D. Cope, "Assessment of the safety of tunnels. Commissioned by the European parliament," Tech. Rep. IP/A/STOA/FWC/2005-28/SC22/29, Published in the European Parliament Web-site under the Rubric "Science and Technology, Options Assessment" (STOA), 2008.
- [12] A. Haack, "Current safety issues in traffic tunnels," *Tunnelling and Underground Space Technology*, vol. 17, no. 2, pp. 117–127, 2002.
- [13] S.-R. Zhang, Z.-L. Ma, and Q. Shi, "Distribution characteristics and countermeasures of traffic accidents in expressway tunnel group," *Journal of Chang'an University*, vol. 27, no. 1, pp. 63–66, 2007.
- [14] Z.-L. Ma, C.-F. Shao, and S.-R. Zhang, "Characteristics of traffic accidents in Chinese freeway tunnels," *Tunnelling and Underground Space Technology*, vol. 24, no. 3, pp. 350–355, 2009.
- [15] F.-B. Lin, C.-W. Chang, P.-Y. Tseng, and C.-W. Su, "Capacity and other traffic characteristics in Taiwan's 12.9-km-long Shea-San tunnel," *Transportation Research Record*, vol. 2130, no. 1, pp. 101–108, 2009.
- [16] Y. Zhang, C. He, D. Wu, and Y. Zeng, "Characteristics and countermeasures of traffic accidents in expressway tunnel," *Journal of Southwest Jiaotong University*, vol. 44, no. 5, pp. 776–781, 2009.
- [17] Z.-L. Ma, C.-F. Shao, and X. Li, "Analysis of factors affecting accident severity in highway tunnels based on logistic model," *Journal of Jilin University*, vol. 40, no. 2, pp. 423–426, 2010.
- [18] G. Y. Jiang and T. Q. Ding, *Traffic Engineering*, National Defense Industry Press, Beijing, China, 2007.
- [19] J. G. Gao, L. Chen, H. Y. Chen, and Q. Chen, "Foreign rural road accidents characteristics," *Journal of China & Foreign Highway*, vol. 29, no. 4, pp. 250–256, 2009.
- [20] "Traffic trip characteristics—spatial and temporal distribution," <http://sh.eastday.com/qtmt/20110309/u1a863059.html>.
- [21] W. H. Zhao, H. X. Liu, M. Guang, and L. D. Shi, "Impact of long tunnel entrance on drivers' psychology and speed," *Highways & Automotive Applications*, no. 3, pp. 48–51, 2011.

## Research Article

# Automated Generation of Traffic Incident Response Plan Based on Case-Based Reasoning and Bayesian Theory

Yongfeng Ma,<sup>1</sup> Wenbo Zhang,<sup>2</sup> Jian Lu,<sup>1</sup> and Li Yuan<sup>3</sup>

<sup>1</sup> Jiangsu Key Laboratory of Urban ITS, Southeast University, 2 Si-Pai Lou, Nanjing, Jiangsu 210096, China

<sup>2</sup> Transportation Engineering and Infrastructure Systems, Department of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907, USA

<sup>3</sup> College of Civil and Transportation Engineering, Hohai University, 1 Xikang Road, Nanjing, Jiangsu 210098, China

Correspondence should be addressed to Yongfeng Ma; mayf@seu.edu.cn

Received 24 September 2013; Accepted 12 December 2013; Published 29 January 2014

Academic Editor: Huimin Niu

Copyright © 2014 Yongfeng Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traffic incident response plan, specifying response agencies and their responsibilities, can guide responders to take actions effectively and timely after traffic incidents. With a reasonable and feasible traffic incident response plan, related agencies will save many losses, such as humans and wealth. In this paper, how to generate traffic incident response plan automatically and specially was solved. Firstly, a well-known and approved method, Case-Based Reasoning (CBR), was introduced. Based on CBR, a detailed case representation and  $R^5$ -cycle of CBR were developed. To enhance the efficiency of case retrieval, which was an important procedure, Bayesian Theory was introduced. To measure the performance of the proposed method, 23 traffic incidents caused by traffic crashes were selected and three indicators, Precision  $P$ , Recall  $R$ , and Indicator  $F$ , were used. Results showed that 20 of 23 cases could be retrieved effectively and accurately. The method is practicable and accurate to generate traffic incident response plans. The method will promote the intelligent generation and management of traffic incident response plans and also make Traffic Incident Management more scientific and effective.

## 1. Introduction

Traffic incidents, randomly occurring spatially and temporally, usually are caused by traffic crashes, disabled vehicles, natural disasters, or leaked hazardous materials, and so on. Due to their unpredictability, complexity, urgency, and potential harmfulness, much negative effects on transportation system are caused, such as much traffic delay, traffic jam, traffic facilities losses, and even secondary traffic crash. And also, there may be many casualties, much economy losses, and ecological failure. In America, statistics from 75 cities in 2000 revealed that 52%–58% of traffic delays were caused by traffic incidents, which would lead to more than \$75 billion economical losses and more than 8.4 billion gallons fuel consuming [1]. In 2008, snow disaster and Wenchuan earthquake hit southern China and led to traffic paralysis in many regions, which greatly influences the rescue work and brings much more losses.

Traffic Incident Management (TIM), as an important component of intelligent transportation system (ITS), is defined as the coordinated and preplanned system. With the help of TIM, human resources and equipment can be managed and distributed timely to reduce the durations and consequences of traffic incidents and then achieve smoother traffic flow. Some successful practices of TIM in developed countries have shown that TIM is an effective and feasible solution to manage traffic incident [2–4]. Traffic incident response plan, as critical module of TIM, is a preplanned scheme to implement emergency rescue and reduce incident-related losses. The plan, which identifies the suitable incident, necessary resources and staffs, emergency measures, and corresponding time limitation, will guide response agencies to make decision and take actions timely and also guarantee quick response to traffic incidents [5]. If the response plans are unreasonable and unfeasible, they will lead to lose the mechanism of coordination and consistency among response

agencies and even make poor accountability [6]. Considering the real-time response and multiagency coordination, generating suitable response plan for target traffic incident quickly and accurately is a key procedure of TIM.

For response plan, there are main three types used in the past few decades (1) Text plan: the text plan, which is an initial type and a static text, provides schemes of responding to potential incidents with words, based on past cases and experiences. (2) Graphic plan: compared with text plan, it takes various types of information to state schemes, such as words, pictures, and videos. (3) Reasoning plan: based on graphic plan, some models are introduced to generate response plan automatically, simulate the implementation of response plan, and then measure performance of the response plan. At present, most agencies manage response plan with the first two types, which are inferior during incident response. Firstly, most response plans are conserved with books or saved in computer simply, which are very difficult to disseminate the accurate plan timely to response agencies. Furthermore, response plans will be revised periodically, which are also difficult with books. Finally, different levels of response plan, which may be applied to a nation, a province or a city, would also influence the incident response. To respond effectively, a much more digital, intelligent, and visual type of response plan should be used, such as reasoning plan. It can generate response plan automatically, query response plan, make decision intelligently, and also show incident response performance visually.

Response plan generation is an important part of reasoning plan, and many researchers have paid attentions on it. Willians classified incident response plan into four parts, which were prior prevention, prior preparation, incident response and recovery. Each part had defined necessary response agencies and their responsibilities [7]. Zografos et al., Fabiano et al., and Chiu and Zheng introduced some independent models, like graphic theory and linear program, into traffic incident response based on advanced data collection and process technologies and showed that these methods were efficient and feasible [8–10]. And also, some Chinese researchers like Dong, Luo et al., and Xiang and Zhang defined response agencies' responsibilities and used Case-Based Reasoning, Rule-Based Reasoning, or dynamic game model to develop traffic incident response plans [11–13]. For these researches, their methods, which were strongly theoretical, enhanced the efficiency of generating response plan in theory. It should be validated further in practice.

The paper mainly focused on developing an algorithm, which could generate appropriate traffic incident response plans timely and automatically. Case-Based Reasoning (CBR), which could obtain an initial plan through retrieving similar incident and then gain the final plan after revision, was introduced to generate response plans. To implement case retrieval effectively, Bayesian Theory was also introduced to search for the similar traffic incident from past traffic incidents. With the designed algorithm, traffic incident response plan could be developed automatically and timely, according to past experiences. It would be helpful to save response time and guide response agencies much more efficiently.

## 2. Methodology

How to automatically generate traffic incident response plans based on prior cases or experiences is the most important problem of the paper. Compared with some other related algorithms, CBR is adopted, which is the process of solving new problems based on similar experiences within the same domain. CBR is a general paradigm for problem solving based on the recall and reuse of specific experiences. And also, it is an approach with incremental and sustained learning, since a new experience is retained each time after iteration [14].

To state the CBR model, the  $R^5$ -CBR cycle, proposed by Finnie and Sun [15], was introduced by adding case repartition into  $R^4$ -CBR cycle. Case repartition could reduce the complexity of case database and facilitate case retrieval. The new  $R^5$ -CBR cycle consisted of case retrieval, case reuse, case revision, case retain, and case repartition, as shown in Figure 1. Based on the cycle, there were the following six steps in a typical CBR:

- (i) traffic incident verification;
- (ii) retrieving the most similar case from case database;
- (iii) reuse of the retrieved case to attempt to solve the current problem;
- (iv) revising the proposed solutions;
- (v) retaining the final solution as a part of database;
- (vi) repartition case database to a satisfactory one.

In the paper, to be combined with Bayesian Theory, CBR was simplified to five steps: traffic incident verification, case representation, case retrieval based on Bayesian Theory, case reuse and revision, and case learning and maintenance.

*2.1. Traffic Incident Verification.* After detecting traffic incidents, some raw information can be obtained from traffic incident detection system. With the help of traffic surveillance or highway patrols, verifying the raw information can be done. To start to generate response plans, the verified raw information should be transmitted to emergency management center standardly and then trigger response plan database. So, a standard table should be designed to report the raw information.

Different factors can cause different traffic incidents, such as disabled vehicles or traffic crashes. So, there are much information to be confirmed, including basic information and representation information. Name of incident-related highway, time, and weather belong to basic information. Traffic incident categories and their corresponding indicators belong to representation information. To record traffic incidents fully and accurately, a table is developed based on specifications in China [16], as shown in Figure 2. The figure is divided into three parts: basic information collection, first-level incident information conformation, and second-level incident information conformation. The incident first-level information are developed to classify traffic incidents into three categories, such as traffic crash, disabled vehicles, and abnormal highway conditions. The incident second-level information is designed to confirm traffic impact and some representation information.

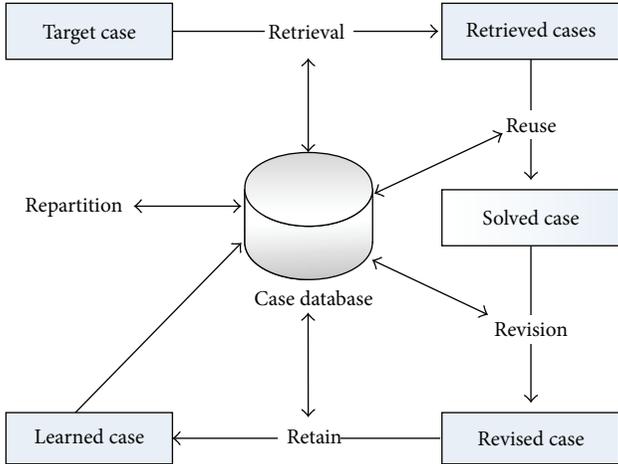


FIGURE 1: The  $R^5$  model of Case-Based Reasoning.

2.2. *Case Representation.* Case representation means to present the cases in the form that the computer can access, deal with, and store. Using ArcGIS Server, a visual web system was developed based on Browser/Server and Java language to manage traffic incident response plans. To develop database for cases, Oracle database was adopted for the real project development. In fact, once the data are stored with the data structure mentioned below, any database is suitable for storing cases. Various case representation methods have been developed, including frame representation and object-oriented representation. Here, frame representation was chosen considering data processing and mathematical algorithm. The main frame is the case of traffic incidents, as shown in Table 1 (taking traffic crashes as example). The response plan of traffic incidents was stored as the subframe, as shown in Table 2.

The main frame contains information obtained from Figure 2. The subframe consists of three slots: traffic incident introduction, countermeasures of traffic incident response, and evaluation of traffic incident response. Five facets, which were agency coordination, traffic management, field management, traffic recovery, and others, were designed for the second slot.

2.3. *Case Retrieval Based on Bayesian Theory.* Case retrieval is the most important step of CBR, which influences greatly the performance of CBR. There are two problems which should be solved, which are case indexing and case retrieval. For case indexing, related cases in database can be obtained quickly given a target case. To develop the case indexing, two principles are proposed: (1) the case indexing should be abstract to some extent to retrieve various cases, but (2) the case indexing should be also a little special, or differences among cases may be vanished. In the paper, traffic incident first-level information, traffic crashes, disabled vehicles, and abnormal highway conditions were selected as case indexing.

Generally, for case retrieval, two goals should be gained: (1) the retrieved cases are as similar as possible with target case and (2) the number of retrieved cases is as few as possible. The prior goal is very important. To meet the two

TABLE 1: Main frame of cases in database.

Frame (cases of traffic crashes)	
ID	Numeric
Time	The time of the traffic incident
Location	The location of the traffic incident
Highway grades	Choice (freeway, others)
Weather	Choice (fine, cloud, rain, snow, fog, and frozen)
Minor wounds	Numeric
Serious injuries	Numeric
Fatalities	Numeric
Stranded persons	Numeric
Fire emergency	Choice (none, slight, and serious)
Turn-over emergency	Choice (none, compact car, and large car)
Number of damaged vehicles	Numeric
Loses of facilities	Choice (none, guardrails, medians, pavement, traffic signs or markings, and others)
Number of occupied lanes	Numeric
Total lanes of the two-way	Numeric
Traffic flow condition	Choice (normal, slow down, congestion, and stopped)
Name of hazardous chemicals	Character
Chemical property	Choice (flammable or explosive, poisonous, and corrosive)
Physical form	Choice (gas, liquid, and solid)
Ranges	Choice (lanes, downwind, and downstream of rivers)
Quantity of leakage	Choice (few, mass)

TABLE 2: Subframe of cases in database.

Subframe (subframe of traffic incident response)	
Traffic incident introduction	
Countermeasures of traffic incident response	
Agency coordination	Scheme of department interaction
Traffic management	Scheme of traffic management
Field management	Scheme of field management
Traffic recovery	Scheme of traffic recovery
Others	Complements
Remarks	Evaluation of traffic incident response

goals, Bayesian Theory, which can predict unknown things with Bayesian Probability based on existing knowledge, is introduced to search for the similar case from case database.

To retrieve similar cases with Bayesian Theory effectively, some problems and mathematical expressions should be identified. Here, there are  $N$  cases, which are independently

**Conformation of traffic incidents in regional highway network**

Location \_\_\_\_\_ Highway stake \_\_\_\_\_ Highway grades  Freeway  Others  
 Signature \_\_\_\_\_ Weather:  Fine  Cloud  Rain  Snow  Fog  Frozen  
 Time \_\_\_\_\_ Number of lanes: \_\_\_\_\_ (one-way)

Incident first-level information

<p style="text-align: center;">Traffic crashes</p> <p>① Are there casualties or losses? <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>② Are there leakages of hazardous chemicals? <input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p style="text-align: center;">Disabled vehicles</p> <p>③ Are there disabled vehicles? <input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p style="text-align: center;">Abnormal highway conditions</p> <p>④ Are there dropped goods or accidental traffic jam? <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>⑤ Are there geological disasters or bad weather? <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>⑥ Are there mass incidents? <input type="checkbox"/> Yes <input type="checkbox"/> No</p>
---	---	--

Incident second-level information

**Traffic crashes**

Number of casualties:

Minor wounds	Serious injuries	Fatalities	Stranded persons

Main serious injuries (multi-choice):

<input type="checkbox"/> Severe trauma <input type="checkbox"/> Open fractures <input type="checkbox"/> Coma <input type="checkbox"/> Severe burns <input type="checkbox"/> Hurt by hazardous chemicals <input type="checkbox"/> Others
---

Fire emergency:  None  Slight  Severe or explosion

Turn-over emergency:  None  Compact car  Large car

**Hazardous chemicals emergency:**

License number of vehicle	Name of hazardous chemical
Chemical property (multi-choice)	Physical form
<input type="checkbox"/> Flammable or explosive <input type="checkbox"/> Poisonous <input type="checkbox"/> Corrosive	<input type="checkbox"/> Gas <input type="checkbox"/> Liquid <input type="checkbox"/> Solid
Ranges (multi-choice)	Quantity of leakage
<input type="checkbox"/> Lanes <input type="checkbox"/> Downwind <input type="checkbox"/> Downstream of rivers	<input type="checkbox"/> Few <input type="checkbox"/> Mass

**Losses of facilities:**

Number of vehicles
Damaged facilities
<input type="checkbox"/> Guardrails <input type="checkbox"/> Medians <input type="checkbox"/> Pavement <input type="checkbox"/> Traffic signs or markings <input type="checkbox"/> Others _____ <input type="checkbox"/> None

**Disabled vehicles**

Number of disabled vehicles	
-----------------------------	--

**Abnormal highway conditions**

<input type="checkbox"/> Dropped goods			
<input type="checkbox"/> Accidental traffic jam	Estimated duration	<input type="checkbox"/> Less than 1 hour <input type="checkbox"/> 1 to 2 hours <input type="checkbox"/> 2 to 4 hours <input type="checkbox"/> More than 4 hours	
<input type="checkbox"/> Geological disasters	Types	<input type="checkbox"/> Surface collapse <input type="checkbox"/> Landslide <input type="checkbox"/> Debris flow <input type="checkbox"/> Earthquake	
<input type="checkbox"/> Fog <input type="checkbox"/> Sandstorms <input type="checkbox"/> Rainstorm	Visibility (meters)	<input type="checkbox"/> 100-200 <input type="checkbox"/> 50-100 <input type="checkbox"/> 30-50 <input type="checkbox"/> Less than 30	
<input type="checkbox"/> Snow or ice	Extent	<input type="checkbox"/> None <input type="checkbox"/> Snow without ice <input type="checkbox"/> Part of section frozen <input type="checkbox"/> All section frozen	
<input type="checkbox"/> Gale	Cross wind force (level)	<input type="checkbox"/> 6 <input type="checkbox"/> 7-8 <input type="checkbox"/> 9-10 <input type="checkbox"/> More than 10	
<input type="checkbox"/> Mass incidents			

**Traffic impact**

Occupied lanes:

Direction	<input type="checkbox"/> Single <input type="checkbox"/> Both-way	Number of occupied lanes	
-----------	--	--------------------------	--

Traffic flow:

<input type="checkbox"/> Normal <input type="checkbox"/> Slow down <input type="checkbox"/> Congestion <input type="checkbox"/> Stopped
--

FIGURE 2: Conformation of Traffic Incidents.

identical distributed random samples, in case database  $D$ . All cases in  $D$  can be represent by  $d_i$  ( $i = 1, 2, \dots, N$ ) and obey an unknown probability distribution  $P$ . There are  $m$  attributes  $X_j$  ( $j = 1, 2, \dots, m$ ) in each case, whose values  $x_j$  are discrete data and shown in (1). Here, some assumptions are represented by  $\Psi$ , and then a model classed with parameters can be represented by  $M$  under  $\Psi$ , in which a model can be represented by  $\Theta$ . Given sample  $D$ , a model  $\Theta(D)$  should be developed to make  $P(d_i | \Theta(D))$  as similar to  $P(d_i)$  as possible:

$$x_j \in \{x_{j1}, x_{j2}, \dots, x_{jn_j}\}, \quad (1)$$

where  $n_j$  is the number of different values of the  $j$ th attributes in case database.

Using Bayesian Theory, to measure the similarity between target case and cases in database, a matching function should be developed, whose values can verify the similarity directly. The more the values of function are, the more similar the retrieved cases are. The expected retrieved case is the maximum value of function among cases database. The definition of function is a critical procedure of case retrieval. Based on Bayesian Theory, the function can be defined initially as conditional probability of each case in database given target case  $u$ , as shown in (2). Unfortunately, the function may only gain nonzero values when cases in database are very identical with target case. Furthermore, the function is in favor of retrieving some cases with larger prior probability, which may not be expected cases. So, (2) is revised, as shown in (3):

$$\text{Sim}(d_i | u) = P(d_i | u, D, \Psi), \quad (2)$$

$$\begin{aligned} \text{Sim}(d_i | u) &= P(u | d_i, D, \Psi) \\ &= \frac{P(d_i | u, D, \Psi)}{P(d_i | D, \Psi)} P(u | D, \Psi) \\ &\propto \frac{P(d_i | u, D, \Psi)}{P(d_i | D, \Psi)}, \end{aligned} \quad (3)$$

where  $\text{Sim}(d_i | u)$  is the similarity between cases in database and target case.

With Bayesian Theory, (4) can be obtained. To use (3) and (4) in practice, proposed assumptions should be in favor of computation of these equations. So, the assumption of Naïve Bayes model was adopted, in which all variables were independent except for a special class variable. In the paper, attribute  $X_m$  is assumed as class variable, and then the probability of case could be computed by (5):

$$\begin{aligned} \text{Sim}(d_i | u) &= P(u | d_i, D, \Psi) \\ &= \int_{\Theta \in M(\Psi)} P(u | \Theta, \Psi) P(\Theta | d_i, D, \Psi) d\Theta, \end{aligned} \quad (4)$$

$$P(d | \Theta) = P(X_m = x_m) \prod_{j=1}^{m-1} P(X_j = x_j | X_m = x_m). \quad (5)$$

Once determining the parameter of model  $\Theta = (\alpha, \Phi)$ , shown in (6), the probability distribution could be also

verified. To simplify the computation, the two parameters were assumed to obey Dirichlet distribution, as shown in (7). With (4)–(8), equation (9) could be computed when the two parameters  $(\alpha, \Phi)$  are independent, as shown in (9):

$$\begin{aligned} \alpha &= (\alpha_1, \alpha_2, \dots, \alpha_k), \\ \alpha_k &= P(X_m = k), \\ \Phi &= (\Phi_{11}, \dots, \Phi_{1(m-1)}, \dots, \Phi_{k1}, \dots, \Phi_{k(m-1)}), \\ \Phi_{sj} &= (\Phi_{sj1}, \dots, \Phi_{sjn_j}), \quad (s = 1, 2, \dots, k), \\ \Phi_{sjl} &= P(X_j = x_{jl} | X_m = x_m), \quad (l = 1, 2, \dots, n_j), \end{aligned} \quad (6)$$

where  $k$  is the number of different values of class variable  $X_m$ . Consider

$$\begin{aligned} (\alpha_1, \alpha_2, \dots, \alpha_k) &\sim \text{Dir}(\alpha | \mu_1, \mu_2, \dots, \mu_k) \\ &= \frac{\Gamma(\mu)}{\prod_{s=1}^k \Gamma(\mu_s)} \prod_{s=1}^k \alpha_s^{\mu_s-1}, \\ (\Phi_{sj1}, \dots, \Phi_{sjn_j}) &\sim \text{Dir}(\Phi_{sjl} | \sigma_{sj1}, \sigma_{sj2}, \dots, \sigma_{sjn_j}) \\ &= \frac{\Gamma(\sigma_{sj})}{\prod_{l=1}^{n_j} \Gamma(\sigma_{sjl})} \prod_{l=1}^{n_j} \Phi_{sjl}^{\sigma_{sjl}-1}, \\ \mu &= \sum_{s=1}^k \mu_s, \quad \mu_s > 0, \\ \sigma_{sj} &= \sum_{l=1}^{n_j} \sigma_{sjl}, \quad \sigma_{sjl} > 0, \\ \text{Sim}(d_i | u) &= \sum_{s=1}^k P(u, X = k | d_i, D, \Psi) \\ &= \sum_{s=1}^k \frac{h_s + \mu_s}{N + \sum_{s=1}^k \mu_s} \prod_j \frac{\sum_{l=1}^{n_j} (f_{sjl} + \sigma_{sjl})}{h_s + \sum_{l=1}^{n_j} \sigma_{sjl}}, \end{aligned} \quad (7)$$

where  $N$  is the number of all cases in database;  $h_s$  is the number of cases whose attribute  $X_m$  are equal to  $k$ ;  $f_{sjl}$  is the number of cases whose attributes  $X_m$  are equal to  $k$  and attributes  $X_j$  are equal to  $x_{jl}$ ;  $\mu_s$  is the hyper-parameter  $\alpha$  of Dirichlet distribution, valued by 1 if without a priori knowledge; and  $\sigma_{sjl}$  is the hyper-parameter  $\Phi$  of Dirichlet distribution, valued by 1 if without priori knowledge.

Finally, a threshold  $\eta$  should be proposed. If  $\text{Sim}(d_i | u)$  is more than or equal to  $\eta$ , then the corresponding case can be retrieved. Here,  $\eta$  is valued by 0.7.

**2.4. Case Reuse and Revision.** In the two steps, two questions should be determined, which part of retrieved response plan can be reused directly in target case and which part of response plan must be revised to be used in target case. The differences between target case and retrieved case are helpful to solve the questions. If the managers are satisfied

TABLE 3: Performance of Case-Based Reasoning and Bayesian Theory.

No.	Current Case	Case retrieval database	Number of retrieved cases ( $n = 0.7$ )	Precision ( $P$ )	Recall ( $R$ )	Indicator ( $F$ )
1	Case 1	Case 2~Case 23	4	68.3%	18.2%	70.6%
2	Case 2	Case 1 and Case 3~23	8	72.4%	36.4%	70.5%
3	Case 3	Case 1~2 and Case 4~23	4	67.2%	18.2%	69.7%
4	Case 4	Case 1~3 and Case 5~23	11	54.7%	50.0%	53.7%
5	Case 5	Case 1~4 and Case 6~23	7	66.5%	31.8%	66.8%
6	Case 6	Case 1~5 and Case 7~23	11	52.3%	50.0%	51.8%
7	Cas 7	Case 1~6 and Case 8~23	12	54.6%	54.5%	52.5%
8	Case 8	Case 1~7 and Case 9~23	6	68.7%	27.3%	69.5%
9	Case 9	Case 1~8 and Case 10~23	8	64.5%	36.4%	64.3%
10	Case 10	Case 1~9 and Case 11~23	3	76.4%	13.6%	78.2%
11	Case 11	Case 1~10 and Case 12~23	8	72.3%	36.4%	70.4%
12	Case 12	Case 1~11 and Case 13~23	7	68.7%	31.8%	68.6%
13	Case 13	Case 1~12 and Case 14~23	9	70.1%	40.9%	67.6%
14	Case 14	Case 1~13 and Case 15~23	3	73.8%	13.6%	76.0%
15	Case 15	Case 1~14 and Case 16~23	2	72.1%	9.1%	75.2%
16	Case 16	Case 1~15 and Case 17~23	3	70.4%	13.6%	73.1%
17	Case 17	Case 1~16 and Case 18~23	7	69.5%	31.8%	69.2%
18	Case 18	Case 1~17 and Case 19~23	2	74.6%	9.1%	77.4%
19	Case 19	Case 1~18 and Case 20~23	4	68.4%	18.2%	70.7%
20	Case 20	Case 1~19 and Case 21~23	5	75.5%	22.7%	75.8%
21	Case 21	Case 1~20 and Case 22~23	3	72.4%	13.6%	74.8%
22	Case 22	Case 1~21 and Case 23	5	73.9%	22.7%	74.6%
23	Case 23	Case 1~Case 22	4	72.3%	18.2%	74.0%

with the retrieved case, it can be reused directly. Otherwise, the retrieved case needs to be revised and improved to deal with its specific condition. Generally, there are four methods to revise case, including substitution, transformation, special-purpose adaption and repair, and derivational replay. In this paper, substitution method was adopted, which could develop new case only through changing some different values in retrieved case.

**2.5. Case Learning and Maintenance.** The CBR can learn new knowledge easily, which guarantees the efficiency and reliability of CBR for a long time. Case learning can be implemented by processing revised case and adding it into case database or replacing corresponding retrieved case. However, if the database for cases is too large, the efficiency and accuracy of case retrieval will be decreased. It is necessary to manage database, such as adding, revising, and deleting some cases. If no suitable cases are retrieved, case made by managers should be added into database after the event. The managers can also revise the retrieved case if necessary. Finally, the managers should delete the useless cases regularly to ensure the efficiency and accuracy.

### 3. Validation

To measure the performance of the method, 23 traffic incidents caused by traffic crashes were selected randomly from

traffic incident database from Traffic Management Bureau, the Ministry of Public Security of China. Due to the space limitations in the paper, the detailed information of 23 cases is not listed. However, the main frame and subframe of crash cases in database are illustrated in Tables 1 and 2. For these 23 cases (case 1~case 23), one case was designed as target case, and the other 22 cases were developed as case retrieval database. Each of these 23 cases could be designed as target case for only once. Therefore, 23 case retrievals were available to measure performance, as shown in Table 3.

Three indicators were introduced to measure performance, including Precision  $P$ , Recall  $R$ , and Indicator  $F$  [17]. The values of these indicators could be calculated with (10). The testing results of case retrieval were shown in Table 3:

$$\begin{aligned}
 P &= \frac{\sum_{i=1}^n N_i}{nN_0}, \\
 R &= \frac{n}{n_A}, \\
 F &= \frac{(\beta^2 + 1)P(1 - R)}{\beta^2P + (1 - R)},
 \end{aligned} \tag{10}$$

where  $n$  is the number of retrieved cases;  $N_i$  is the number of the  $i$ th retrieved case's attributes whose value is equal to that of current case;  $N_0$  is the number of attributes for each

case;  $n_A$  is the number of all related cases in database;  $\beta$  is the significance coefficient between Precision and Recall, 0.5 used, which means Precision's significance is twice more than Recall's.

#### 4. Conclusion

Incident response plan, as an important component of Traffic Incident Management, provides schemes to respond to traffic incidents and identifies response agencies. The paper mainly focused on the automated generation of traffic incident response plan, which was a function of reasoning plan. To cope with the problem, an effective model, known as Case-Based Reasoning, was introduced to design, retrieve, revise, and manage cases. To enhance the efficient and accuracy of case retrieval, Bayesian Theory was also introduced to develop a function to measure the similarity between cases in database and target case.

Based on the testing dataset containing 23 traffic incidents cases, the method was verified to be feasible and effective. The validation results showed that most of target cases could be retrieved effectively using the method, except for case 4, case 6, and case 7. The 23 cases in tested database were insufficient to retrieve all target cases effectively and accurately. A better performance could be achieved by increasing the amount of cases in the database to an appropriate number. With the proposed method, traffic response plan could be generated automatically and timely. And also, the method could collect prior successful experiences and manage them. These were all helpful for incident managers and authorities to make decisions and coordinate with other agencies.

As for future work, differences among attributes should be developed more completely. Using Bayesian Theory, all attributes were equally important. However, in fact, there were some differences among them. Much more attention might be paid on the attribute of causalities than that of damaged vehicles, when generating traffic incident response plans. So during case retrieval, weights for each attribute could be introduced to enhance the performance of case retrieval.

#### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

#### Acknowledgments

This study is funded by National Natural Science Foundation (no. 51208100). The authors also would like to thank the graduate research assistants from the School of Transportation, Southeast University for their assistance in data processing.

#### References

- [1] V. Pearce and S. Subramaniam, *Incident Management: Detection, Verification and Traffic Management, Field Operational*

- Test Cross-Cutting Study*, U.S. Department of Transportation, Washington, DC, USA, 1998.
- [2] C. Rochon, "Performance evaluation and benefit analysis for CHART-coordinated highways action response team," Year 2005 Final Report, University of Maryland, College Park, Md, USA, 2006.
- [3] L. Hagen, H. Zhou, and H. Singh, *Road Ranger Benefit-Cost Analysis*, University of South Florida, Tampa, Fla, USA, 2005.
- [4] H. HenkR, M. E. Molina, and P. L. Irwin, "Before-and-after analysis of the San Antonio transGuide system," in *Proceedings of the 76th Annual Meeting of the Transportation Research Board*, CD-ROM, Washington, DC, USA, January 1997.
- [5] Office of Emergency Management, *General Office of the State Council of the People's Republic of China. Overall Emergency Response Plan for National Public Incident*, China Legal Publishing House, Beijing, China, 2006.
- [6] Q. Liu, H. P. Lu, Y. B. Zhang et al., "Characteristic analysis and countermeasure study on road traffic accidents in China," *China Safety Science Journal*, vol. 16, no. 6, pp. 123–128, 2006.
- [7] G. Williams, S. Batho, and L. Russell, "Responding to urban crisis: the emergency planning response to the bombing of Manchester city centre," *Cities*, vol. 17, no. 4, pp. 293–304, 2000.
- [8] K. G. Zografos, K. N. Androutsopoulos, and G. M. Vasilakis, "A real-time decision support system for roadway network incident response logistics," *Transportation Research C*, vol. 10, no. 1, pp. 1–18, 2002.
- [9] B. Fabiano, F. Currò, A. P. Reverberi, and R. Pastorino, "Dangerous good transportation by road: from risk analysis to emergency planning," *Journal of Loss Prevention in the Process Industries*, vol. 18, no. 4–6, pp. 403–413, 2005.
- [10] Y.-C. Chiu and H. Zheng, "Real-time mobilization decisions for multi-priority emergency response resources and evacuation groups: model formulation and solution," *Transportation Research E*, vol. 43, no. 6, pp. 710–736, 2007.
- [11] J. R. Dong, *Research on Constitution Technology of National Economy Mobilization Preparedness*, Huazhong University of Science and Technology, Wuhan, China, 2007.
- [12] J. W. Luo, Z. P. Shi, Q. He, Z. Shi et al., "A quick emergency response plan generation system combining CBR and RBR," *Journal of Computer Research and Development*, vol. 44, no. 4, pp. 660–666, 2007.
- [13] H. Y. Xiang and Z. Zhang, "Model of emergency plan for expressway accident rescue," *Journal of Dalian Jiaotong University*, vol. 30, no. 3, pp. 8–11, 2009.
- [14] A. G. de Silva Garza and M. L. Maher, "An evolutionary approach to case adaption, case-based reasoning research and applications," in *Proceedings of the 3rd International Conference on Case-Based Reasoning*, Munich, Germany, July 1999.
- [15] G. Finnie and Z. Sun, "R<sup>5</sup> model for case-based reasoning," *Knowledge-Based Systems*, vol. 16, no. 1, pp. 59–65, 2003.
- [16] Standardization administration of the people's republic of China, *Road Traffic Information Collection—Incident Information*, China Standards Press, Beijing, China, 2006.
- [17] S. Buttcher, L. A. Charles Clarke, and V. Gordon Cormack, *Information Retrieval: Implementing and Evaluating Search Engines*, The MIT Press, Cambridge, Mass, USA, 2010.

## Research Article

# Optimization Based High-Speed Railway Train Rescheduling with Speed Restriction

Li Wang,<sup>1,2</sup> Wenting Mo,<sup>3</sup> Yong Qin,<sup>1</sup> Fei Dou,<sup>2</sup> and Limin Jia<sup>1</sup>

<sup>1</sup> State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup> School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

<sup>3</sup> IBM China Research Lab, 2F Diamond A, Zhong Guan Cun Software Park, Beijing 100193, China

Correspondence should be addressed to Limin Jia; jialm@vip.sina.com

Received 16 August 2013; Revised 17 October 2013; Accepted 6 November 2013; Published 12 January 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Li Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A decision support framework with four components is proposed for high-speed railway timetable rescheduling in case of speed restriction. The first module provides the speed restriction information. The capacity evaluation module is used to evaluate whether the capacity can fulfill the demand before rescheduling timetable based on deduction factor method. The bilayer rescheduling module is the key of the decision support framework. In the bilayer rescheduling module, the upper-layer objective is to make an optimal rerouting plan with selected rerouting actions. Given a specific rerouting plan, the lower-layer focuses on minimizing the total delay as well as the number of seriously impacted trains. The result assessment module is designed to invoke the rescheduling model iteratively with different settings. There are three prominent features of the framework, such as realized interaction with dispatchers, emphasized passengers' satisfaction, and reduced computation complexity with a bilayer modeling approach. The proposed rescheduling model is simulated on the busiest part of Beijing to Shanghai high-speed railway in China. The case study shows the significance of rerouting strategy and utilization of the railway network capacity in case of speed restriction.

## 1. Introduction

China is extensively developing the infrastructure of high-speed railway. The target is to cover its major economic areas with a high-speed railway network, which consists of four horizontal and four vertical lines, in the following several years. The network scale is much larger than any other existing ones in the world. In the network, Beijing-Shanghai high-speed line connects Beijing (the capital of China) and Shanghai (the biggest economic centre of China), and it goes through Yangtze River Delta region (the best developed area in China). Thus, people describe it as the North-South Aorta of China. As designed, trains of different speeds will be operated in a mixed way on the network with a minimum headway of 3 minutes. In other words, it has the characteristics of high train speed, high train frequency, and mixed train speed (HHM).

A basic train timetable essentially provides the arrival time of trains at stations and the corresponding departure time from the stations. Based on the timetable, the dwell

time of trains at the stations (which is 0 if they do not stop) and the departure order of trains from the trains can be derived. During daily operation, disturbances to the timetable are inevitable, which may be caused by emergencies like natural disasters, accidents, or maintenance requests. In case of the disturbance, dispatchers must adopt proper emergency management strategies to ensure operation safety. Consequently, they need to reschedule the timetable to avoid conflict and recover to the original timetable. Drivers will then adjust the train speed and dwell time according to the new arrival and departure time given in the rescheduled timetable.

In the context of transportation operation, the timetable rescheduling problem is famous of its NP hardness. What is worse, it is very hard to find a generalized rescheduling algorithm suitable for all kinds of emergency cases because of various railway infrastructure and various emergency management actions. Except adjusting arrival/departure time, all the actions taken for rescheduling are called rerouting actions

in this paper, including cancelling trains, merging trains, and making a detour.

Setting speed restriction is one of the most common strategies used by dispatchers facing emergency cases. With fixed departure order of trains at all the stations and unchanged time interval between consecutive departures, the throughput of a railway section over a period of time will decrease in case of speed restriction. In other words, trains may be delayed or even cancelled. The impacted number of trains is decided by the size of the restricted section, the time period in restriction, and the restricted speed. On June 20, 2010, the southern part of the existing Beijing-Shanghai line experienced speed restriction for heavy rain storm. As announced, 18 trains departing from Shanghai were cancelled, thereby more than 20,000 passengers being affected. Also, the rest in-service trains were more or less delayed. Considering the high train speed and high train frequency of high-speed lines, the impact of speed restriction would be more serious than on existing lines (in this paper, the existing nonhigh speed railway line is called existing line for short). On the other hand, high-speed lines are more passenger-oriented than existing lines, where punctuality is extraordinarily important. Thus, it is significant to build a responsive and effective decision support mechanism to handle the timetable rescheduling in case of speed restriction for high-speed lines. However, there are two challenges to achieve this goal. One is the interaction between the rescheduling algorithm and the dispatchers. Since the dispatchers have rich experience, their advice and decision on the rescheduling strategies are highly valuable. The other one is the HHM characteristics of the high-speed network.

In this paper, we investigate the design and implementation of a decision support mechanism for timetable rescheduling in case of speed restriction as mentioned above. Note that the three prominent features of this mechanism are realized interaction with dispatchers, emphasized passengers' satisfaction, and reduced computation complexity with a bilayer modeling approach. In the rest of the paper, related research work is discussed in the section of The Literature Review. The next section describes the decision support mechanism in detail, including the mathematical modeling. A case study on the busiest part of Beijing-Shanghai high-speed line is given in the section of Case Study. The last section concludes the whole paper and gives directions for future research.

## 2. The Literature Review

The timetable rescheduling problem has been widely investigated by researchers. Some related works are reviewed, although they are quite different from our work from model creation to solving method.

Some researchers take the rerouting, cancellation, and other complicated strategies into consideration. Reference [1] developed a detailed model based on the identification of possible route conflicts with high accuracy, using the blocking time theory, when the objective is to minimize additional running times. In presence of disturbances, an algorithm detects the infeasible train routes and solves

each conflict locally based on train priorities. Reference [2] regarded the train rescheduling problem as a constraint optimization problem, using passengers' dissatisfaction as the objective criterion. Then an efficient algorithm combining PERT and metaheuristics has been addressed. Experiments show that it works quite fast and it supports versatile methods of rescheduling including cancellation, change of train-set operation schedule, and change of tracks. D'Ariano et al. have developed a real-time dispatching system, called ROMA (railway traffic optimization by means of alternative graphs), to automatically recover disturbances. ROMA is able to automatically control traffic, evaluating the detailed effects of train [3] and local rerouting actions [4], while taking into account minimum distance headways between consecutive trains and the corresponding variability of train dynamics [5–7]. In order to handle large time horizons within a linear increase of computation time, the temporal decomposition approaches which decompose a long time horizon into tractable intervals to be solved in cascade with the objective of improving punctuality have been proposed in [8].

To speed up trains rescheduling computation time, some decomposition or multilayer methods are adopted. Reference [8] proposed the temporal decomposition approach that decompose a long time horizon into tractable intervals. Reference [9] proposed a bilayer optimization model within a simulation framework to deal with the high-speed railway (HSR) line planning problem. It can reduce computation complexity, and an optimal set of stop-schedules can always be generated with less calculation time. Reference [10] presented an optimization method for fast construction of time tables which can be used for dispatching or long term operation planning. This method contains two steps: constructing and iterative improving. Constructing step uses branch-and-bound method to construct the first solution taking into account only restricted amount of possible decisions. Improving step adopts the genetic algorithm as an iterative improving method. These several levels of optimization method can help reduce the calculation effort.

Furthermore, [11, 12] addressed the problem of solving conflicts in railway traffic due to disturbances. The problem is formulated as a problem of rescheduling meets and overtakes of trains and has been dealt with in a two-level process. The upper level handles the order of meets and overtakes of trains on the track sections while the lower level determines the start and end times for each train and the sections it will occupy. Reference [13] dealt with analyzing dispatchers' decision process in intertrain conflict resolutions and developing a heuristic algorithm for rescheduling trains by modifying existing meet/pass plans in conflicting situations in a single-track railway. A system's approach is used in construction of the heuristic algorithm, which is based on inter-train conflict management. Reference [14] introduced an optimization model based on improved symmetric tolerance approach, which could reschedule the timetable timely due to unexpected interference with little cost and risk. Reference [15] investigated the train re-scheduling problem by optimization and simulation in order to obtain an exact or approximate solution. The model aims at maximizing the number of passengers transported and is solved by heuristic procedure

based on backtracking algorithm. The model and DSS are implemented in Asturias of the Spanish National Railway Company in 1998. Reference [16] proposed a model and a solution method for the dispatching of trains on a single-track line. Their model mainly addresses the operational problem of dispatching trains in real time but can also serve at the strategic level to evaluate the impacts of timetable or infrastructure changes on train arrival times and train delays. The formulation is a nonlinear mixed integer program that incorporates lower and upper limits on train velocities for each train on each segment. The objective function only seeks to minimize a combination of total train tardiness and fuel consumption. Based on the driver's safety approaching behavior and pedestrian safety crossing behavior, Wang et al. [17] and Guo et al. [18] proposed an intelligent driving shaping model with multiruled decision-making mechanism in the urban traffic environment.

Despite the effort devoted to solve the sophisticated rescheduling problem, neither of them proposes the closed loop bilayer feedback approach. This approach is able to simulate the scheduling process of dispatchers. Also it permits the dispatchers adding their experiences into the rescheduling process so as to improve the usability of the real-time rescheduling.

### 3. Decision Support Mechanism (DSM) for Timetable Rescheduling

Note that the railway lines considered in this paper are all double-track passenger lines with automatic blocking.

In order to handle daily operation tasks, there is usually a train operation system for dispatchers. In emergencies, dispatchers issue action commands through this system to signal system. There are 4 components in the proposed DSM as shown in Figure 1.

The speed restriction action as the first part is the input coming from the train operation system. It contains the following information.

- (i) Restricted section is a part of the line where speed of trains is restricted.
- (ii) Restricted period is the start and end times of a speed restriction.
- (iii) Applied train type is different trains may be restricted to different speeds according to their type.
- (iv) Speed limitation is the maximum speed a train that can run at.

Given the speed restriction information, the line capacity evaluation module is used to evaluate whether the capacity can fulfill all the trains scheduled in the original timetable in limited service hour. Then, the speed and capacity information are transferred to the third module. In the bilayer rescheduling model, the upper-layer objective is to make an optimal rerouting plan with selected rerouting actions. Given a specific rerouting plan, the second layer focuses on minimizing the total delay as well as the number of seriously impacted trains. Thereafter, if the rescheduling result is

acceptable, the new timetable generated from the DSM will be fed back into the train operation system. Otherwise, the bilayer rescheduling will be executed in a loop till the new timetable is acceptable, so that it can be disseminated to drivers.

*3.1. Line Capacity Evaluation.* As mentioned in the Introduction section, the line throughput may drop dramatically during the restricted period. As a result, it may not be able to fulfill all the trains scheduled in the original timetable in limited service hour. If the capacity was not enough for the demand, the result of timetable rescheduling would be meaningless. Thus, a module is used to evaluate whether the capacity can fulfill the demand before rescheduling timetable. With the evaluation result, the dispatcher can make a decision on train canceling according to rules and their experience.

During the past several years, analytical [19–21] and simulation methods [22] are used to calculate the capacity of railway lines (either single-track or double-track). Among those methods, [23] is widely used in Europe, which can be categorized as a simulation method. It stated that railway capacity depends on both infrastructure and the timetable. Therefore, the capacity calculation according to UIC 406 requires an actual timetable. Some timetabling softwares such as RailSys have already implemented such calculation methods. However, this approach may cost long computation time and may not obtain a feasible timetable.

In this paper, we use deduction factor method [24], which is popular in recent years in China railway, to calculate the line capacity. The deduction factor method is an analytical method for calculating the capacity of high-speed rail line with various types of trains. This method considers the capacity occupancy of different trains and normalizes them into one kind of trains in order to obtain the theoretical value of capacity.

If there is only one kind of trains without stopovers on high-speed rail line, capacity of high-speed rail line with high-speed trains without stopovers can be obtained by

$$n_{\max} = \frac{1440 - t_{\text{maintenance}} - t_{\text{inefficacy}}}{I}, \quad (1)$$

where 1440 is the total number of minutes in a day and  $t_{\text{maintenance}}$  denotes the maintenance window time. In this paper,  $t_{\text{inefficacy}}$  is the capacity, which cannot be used. For instance, we have to ensure that all the trains arrive to destination before the maintenance operation starts. Hence, given the speed of trains, the last train must depart from the first station before a specific time, after which we cannot send more trains. In other words, some capacity is wasted. This is the concept behind the calculation of capacity inefficacy. In addition,  $I$  is the headway between consecutive trains. This parameter is the so-called safety distance, which is determined by the number of blocks between each consecutive train. So it can be obtained by train speed and blocking distance.

On the contrary, if there are various trains with different speed on the rail line, (1) cannot precisely obtain the line capacity, as different kinds of trains may take up different

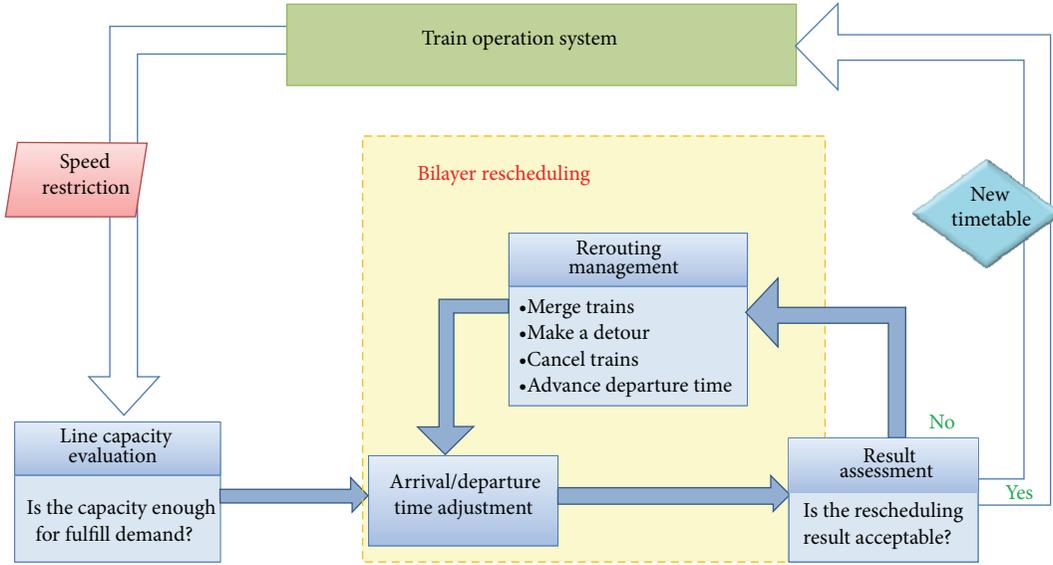


FIGURE 1: Architecture of the decision support mechanism.

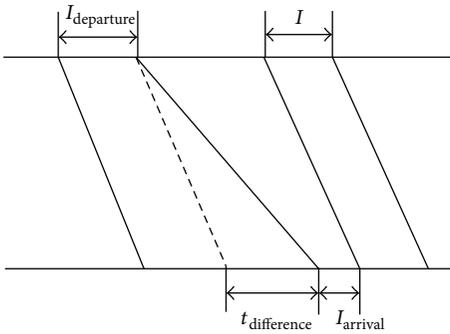


FIGURE 2: Deduction factor for single-quasihigh speed trains without stopovers with regard to high-speed trains without stopovers.

capacity. For example, as shown in Figure 2, quasihigh-speed trains will take up more “area” in the timetable. The deduction factor for single quasihigh-speed trains without stopovers with regard to high-speed trains without stopovers can be obtained by

$$\varepsilon_{\text{high}}^{\text{nostop,quasi}} = \frac{I_{\text{departure}} + t_{\text{difference}} + I_{\text{arrival}} - I}{I}. \quad (2)$$

Different deduction factors need to be calculated according to train operation mode, which are beyond the scope of this paper. After obtaining all kinds of deduction factors, average deduction factor can be obtained according to the ratio of different situations. Then the normalized line capacity can be calculated.

In China, there are two types of trains on most high-speed lines (high-speed trains and quasi-high-speed trains). In this situation, two kinds of deduction factors are used for capacity calculation. One arises from high-speed trains with stopovers with regard to high-speed trains without stopovers. The other

one arises from the speed difference of two types of trains. A detailed example will be given in the Case Study section.

**3.2. Bilayer Rescheduling.** As shown in Figure 1, the rescheduling consists of three components that are executed in a loop. From the optimization perspective, the rescheduling problem is decomposed into two layers. The first layer decides the rerouted trains and the second layer adjusts the arrival and departure times of the rest of the trains. The detailed descriptions of the components are given below.

**3.3. Rerouting Management.** The first step in rescheduling is to decide the rerouting actions to avoid extensive delay resulted from decreased capacity.

If the capacity is not enough, it will be necessary to reduce the demand. In this paper, the following rerouting actions are considered for capacity releasing.

- (i) Merging trains means to merge two trains into one, subject to two constraints. First, only trains of certain type can be merged to prevent the merged train from exceeding the maximum train length. Second, train  $A$  and train  $B$  can be merged only if  $SS_A \subseteq SS_B$  or  $SS_B \subseteq SS_A$ , where  $SS_A$  and  $SS_B$  denote the stopping station set of  $A$  and  $B$ , respectively. Although one train number will disappear after merging, the passengers who bought the tickets of the corresponding train will be noticed to get onboard to the merged train and are guaranteed to be able to get off at their destinations by the second merging constraint. Thus, the impact on the passengers of this rerouting action may be ignored.
- (ii) Making a detour means to drive trains off the high-speed line and use the intercity line or existing line when residual capacity is available.

- (iii) Cancelling means to cancel trains. When a train is cancelled, not only the tickets must be refunded but also the reputation would be damaged. So, this rerouting action is of the lowest priority.

Note that the calculated capacity is the upper bound of the real capacity. Theoretically, if the capacity is larger than the demand, all the trains can travel from the origin to the destination in the service hour. Nonetheless, the capacity may not be fully utilized due to sparse departure from the origin station. Thus, advancing departure time is another rerouting action, which can be taken when the capacity is not fully utilized.

The optimization at this layer aims at finding an optimal rerouting plan with the minimum cost. The cost is associated with the impacts of the above rerouting actions on passengers, including the refundable paid by railway company and the extra transfer taken by passengers. This reflects the passenger-oriented principle in the high-speed railway.

**3.4. Arrival/Departure Time Adjustment.** Given the rerouting plan, the trains remaining to run on the high-speed line are known. Thereafter, the optimization objective is to minimize the weighted delay of these trains, where the weights correspond to their priorities. The objective can be achieved by adjusting the arrival/departure time to/from stations. In practice, the speed of trains running between two consecutive stations, the dwell time of trains at stations, and the departure order of trains from stations may need to be changed according to the arrival/departure time adjustment.

The advantages of bilayer decomposition can be interpreted from different angles. From the system perspective, it facilitates the interaction with dispatchers. They are enabled to explicitly control the rerouting actions. From the problem solving perspective, it effectively reduces the complexity of optimization. Long computation time is usually a curse for rescheduling problems because of the large number of decision variables, which prevents the application of optimization-based automatic timetable rescheduling algorithms being used in real world.

**3.5. Decision Tree-Based Result Assessment.** A decision tree is mined from an evaluation rule set for rescheduling result assessment purpose. A rule defines the conditions for accepting or rejecting a rescheduled timetable. The rule set is generated from the dispatchers' experience on the acceptability of a rescheduled timetable. Basically, extensive delay and long delay of single trains are not acceptable to passengers in high-speed lines. Table 1 gives an example of the evaluation rules, where  $\theta_1 = 30$  mins and  $\theta_2 = 60$  mins.

Overlapping or redundancy may exist in the rule set, as the rules are generated from experience. Greedy algorithm can be used to search for the smallest decision tree based on a given rule set [23]. In a decision tree, each node is a test for an attribute, each branch corresponds to test results, and each leaf assigns a decision. For illustration purpose, Figure 3 shows the smallest decision tree obtained from the rules given in Table 1.

## 4. Mathematical Modeling of the Bilayer Rescheduling

Mixed integer linear programming (MILP) is used to model the proposed bilayer rescheduling.

**4.1. Input Data.** The numerical inputs are described as follows:

$inis_p, inie_p$ : initial start and end times of interval  $p$  according to original timetable;

$z_p$ : the minimum size of interval  $p$ . If  $p$  is the possession of a station,  $d_p$  stands for the minimum dwell time of the train associated with  $p$ . Otherwise, it stands for the minimum running time of the train;

$f_k$ : the minimum separation time of two intervals on station  $k$ , that is, end time of an interval and start time of the subsequent interval;

$hw_k$ : the headway of section  $k$ , that is, start time of an interval and start time of the subsequent interval;

$g_i$ : the length of train  $i$ . Only if  $g_i = G$ , train  $i$  can be merged.  $G$  is a constant input;

$c_i^{\text{cancel}}$ ,  $c_i^{\text{detour}}$ , and  $c_{ij}^{\text{merge}}$ : the cost of canceling train  $i$ , detouring train  $i$ , and merging train  $i$  to train  $j$ , respectively;

$c_i^{\text{delay}}$ : the cost per time unit delay for train  $i$ ;

$w_i$ : delay tolerance for train  $i$ ;

$$h_p \begin{cases} 1, & \text{if interval } p \text{ is a planned stop,} \\ & \text{where } p \in \text{Itrv} \\ 0, & \text{otherwise;} \end{cases} \quad (3)$$

$\Delta C$ : the total residual capacity of alternative lines;

$\delta$ : the minimal number of trains to be applying rerouting actions, which is obtained from line capacity evaluation module;

$M$ : a very big integer, for example, 100000.

Some sets are defined in Table 2 for modeling. Note that the segment between two stations is called "section", while the station segment is called "station" for short.

### 4.2. Model of Rerouting Management Layer

#### 4.2.1. Decision Variables

$$q_i = \begin{cases} 1, & \text{if train } i \text{ is canceled, where } i \in \text{Trn} \\ 0, & \text{otherwise,} \end{cases}$$

$$m_{ij} = \begin{cases} 1, & \text{if train } i \text{ is merged to train } j, \\ & \text{where } i, j \in \text{Trn}, g_i = g_j = G \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

TABLE 1: Example of evaluation rules.

Percentage of trains being delayed more than $\theta_1$	Percentage of trains being delayed more than $\theta_2$	Number of delayed highpriority trains	Acceptable (outcome)
$\leq 80\%$	$\geq 50\%$	$\geq 8$	No
$\leq 20\%$	—	—	Yes
$\leq 50\%$	—	$\leq 15$	Yes
—	$\leq 50\%$	$\leq 15$	Yes
$\vdots$			

TABLE 2: Sets defined for modeling.

Set name	Description	Size	Index of set
Trn	All the trains	$T$	$i$
Seg	All the segments	$S$	$k$
Stn $\subseteq$ Seg	All the stations	$N$	—
Sec $\subseteq$ Seg	All the sections	$S-N$	—
Itv	All the intervals where an interval is the possession of a segment by a train with specified start and end times	—	$p$
$\text{Itv}_i \subseteq \text{Itv}$	The ordered set of intervals for train $i$ according to the original timetable	—	—
$\text{Itv}_i^{\text{stn}} \subseteq \text{Itv}_i$	The ordered set of intervals associated with stopping at stations for trains $i$	—	—
$\text{Itv}_k \subseteq \text{Itv}$	The ordered set of intervals for segment $k$ according to the original timetable	—	—
$\text{Trk}_k$	All the tracks of station $k$	$L_k$	$l$

Remarks: (1) for the ease of understanding, we use  $p$  as the index for every interval-related set (i.e., Itv,  $\text{Itv}_i$ ,  $\text{Itv}_i^{\text{stn}}$ , and  $\text{Itv}_k$ ). In each of the interval-related sets,  $p+1$  indicates the first subsequent interval of  $p$ , and  $\bar{p}$  indicates all the subsequent intervals of  $p$ .

Note that if train  $i$  is merged to train  $j$ , the train number of  $i$  will no longer exist:

$$o_i = \begin{cases} 1, & \text{if train } i \text{ makes a detour with existing} \\ & \text{line, where } i \in \text{Trn} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

#### 4.2.2. Objective Functions

(i) To minimize the rerouting cost,

$$\text{Minimize } \sum_{i \in \text{Trn}} \left( c_i^{\text{cancel}} q_i + c_i^{\text{detour}} o_i + c_{ij}^{\text{merge}} \sum_{j \in \text{Trn}} m_{ij} \right). \quad (6)$$

#### 4.2.3. Constraints

(ii) Cancelled trains cannot make detour.

$$q_i + o_i \leq 1, \quad i \in \text{Trn}. \quad (7)$$

(iii) If a train is eligible to be merged, it can be merged to (with) at most one other train. The merged train cannot be cancelled:

$$\sum_{i=j} m_{ij} = 0, \quad i, j \in \text{Trn},$$

$$\sum_{j \in \text{Trn}} m_{ij} + q_i \leq 1, \quad i \in \text{Trn},$$

$$\sum_{i \in \text{Trn}} m_{ij} + q_j \leq 1, \quad j \in \text{Trn},$$

$$m_{ij} + m_{ji} \leq 1, \quad i, j \in \text{Trn}.$$

(8)

(iv) The number of detoured trains cannot exceed the residual capacity of alternative lines:

$$\sum_{i \in \text{Trn}} o_i \leq \Delta C. \quad (9)$$

(v) The minimal number of the trains that being applied rerouting actions:

$$\sum_{i \in \text{Trn}} \sum_{j \in \text{Trn}} m_{ij} + \sum_{i \in \text{Trn}} q_i + \sum_{i \in \text{Trn}} o_i \geq \delta. \quad (10)$$

#### 4.3. Model of Arrival/Departure Time Adjustment Layer

4.3.1. *The Decision Variables.* The decision variable are described as follows:

$s_p, e_p$ : start time and end time of interval  $p$ ;

$d_p$ : delay of interval  $p$ , which is defined as the difference between the departure time after adjustment and the planned departure time in the original timetable:

$$b_i = \begin{cases} 1, & \text{if train } i \text{ reaches its final considered stop with} \\ & \text{a delay larger than } w_i, \text{ where } i \in \text{Trn} \\ 0, & \text{otherwise,} \end{cases}$$

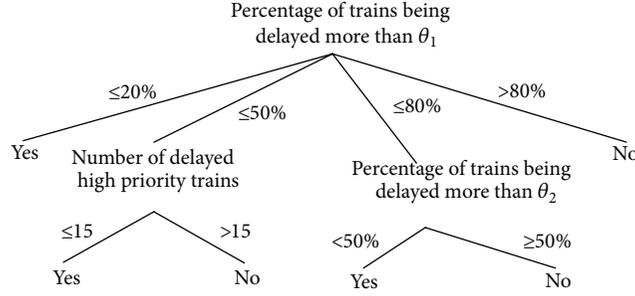


FIGURE 3: Illustration of the smallest decision tree obtained from Table 1.

$$\eta_{pl} = \begin{cases} 1, & \text{if interval } p \text{ uses track } l, \text{ where } p \in \text{Itv}_k, \\ & l \in \text{Trk}_k \text{ and } k \in \text{Stn} \\ 0, & \text{otherwise,} \end{cases}$$

$$\alpha_{p\tilde{p}} = \begin{cases} 1, & \text{if interval } p \text{ occurs before } \tilde{p} \text{ as in the} \\ & \text{initial timetable, where } p, \tilde{p} \in \text{Itv}_k, k \in \text{Seg} \\ 0, & \text{otherwise,} \end{cases}$$

$$\beta_{p\tilde{p}} = \begin{cases} 1, & \text{if interval } p \text{ is changed to occur after } \tilde{p}, \\ & \text{where } p, \tilde{p} \in \text{Itv}_k, k \in \text{Seg} \\ 0, & \text{otherwise.} \end{cases}$$

(11)

#### 4.3.2. Objective Functions

(i) To minimize the delay cost,

$$\text{Minimize } \sum_{i \in \text{Trn}} \left( c_i^{\text{delay}} \cdot \sum_{p \in \text{Itv}_i^{\text{stn}}} d_p \right). \quad (12)$$

(ii) To minimize the number of trains exceeding delay tolerance,

$$\text{Minimize } \sum_{i \in \text{Trn}} b_i. \quad (13)$$

#### 4.3.3. Constraints

(i) Interval restrictions are as follows:

$$e_p \geq s_p + z_p, \quad p \in \text{Itv}. \quad (14)$$

The real departure time cannot be earlier than the original departure time:

$$\begin{aligned} e_p &\geq \text{inie}_p, \quad p \in \text{Itv}, \quad h_p = 1, \\ e_p - \text{inie}_p &= d_p, \quad p \in \text{Itv}. \end{aligned} \quad (15)$$

(ii) Track restrictions

$$\sum_{l \in \text{Trk}_k} \eta_{pl} = 1, \quad p \in \text{Itv}_k, \quad k \in \text{Stn}. \quad (16)$$

(iii) Connectivity restrictions are as follows:

$$e_p = s_{p+1}, \quad p \in \text{Itv}_i, \quad p \neq \text{last}(\text{Itv}_i), \quad i \in \text{Trn}. \quad (17)$$

 For each station, if two intervals use the same track, at least one of  $\alpha$  and  $\beta$  is forced to be 1:

$$\eta_{pl} + \eta_{\tilde{p}l} - 1 \leq \alpha_{p\tilde{p}} + \beta_{p\tilde{p}}, \quad (18)$$

$$p, \tilde{p} \in \text{Itv}_k, \quad p < \tilde{p}, \quad l \in \text{Trk}_k, \quad k \in \text{Stn}.$$

One train can enter a station after the preceding train, which uses the same track, leaves it:

$$s_{\tilde{p}} - e_p \geq f_k \alpha_{p\tilde{p}} - M(1 - \alpha_{p\tilde{p}}), \quad p, \tilde{p} \in \text{Itv}_k, \quad k \in \text{Stn}$$

$$s_p - e_{\tilde{p}} \geq f_k \beta_{p\tilde{p}} - M(1 - \beta_{p\tilde{p}}), \quad p, \tilde{p} \in \text{Itv}_k, \quad k \in \text{Stn}. \quad (19)$$

 For each section, at least one of  $\alpha$  and  $\beta$  is forced to be 1 because there is only one track:

$$\alpha_{p\tilde{p}} + \beta_{p\tilde{p}} = 1, \quad p, \tilde{p} \in \text{Itv}_k, \quad k \in \text{Sec}. \quad (20)$$

Headways for each section are as follows:

$$s_{\tilde{p}} - s_p \geq hw_k \alpha_{p\tilde{p}} - M(1 - \alpha_{p\tilde{p}}), \quad p, \tilde{p} \in \text{Itv}_k, \quad k \in \text{Sec}$$

$$s_p - s_{\tilde{p}} \geq hw_k \beta_{p\tilde{p}} - M(1 - \beta_{p\tilde{p}}), \quad p, \tilde{p} \in \text{Itv}_k, \quad k \in \text{Sec}. \quad (21)$$

(iv) Auxiliary restrictions are as follows:

$$d_{\text{last}(\text{Itv}_i)} - w_i \leq Mb_i. \quad (22)$$

## 5. Case Study

The proposed DSM is simulated on the busiest part of Beijing-Shanghai high-speed line, between Nanjing (Ning for short) and Shanghai (Hu for short). In the rest of this paper, ‘‘Hu-Ning Part’’ is used to represent this part of the Beijing-Shanghai high-speed line. As shown in Figure 5, besides Beijing-Shanghai high-speed line, there are two other lines connecting Hu and Ning, which are Hu-Ning inter-city high-speed line and existing Beijing-Shanghai line. Since all of

them are double-track lines, we only consider the direction from Ning to Hu without loss of generality. All the MILP models are solved by IBM ILOG CPLEX 12.2.

As shown in Figure 4, there are seven stations on the Hu-Ning Part, Nanjing South, Zhenjiang West, Changzhou North, Wuxi East, Suzhou North, Kunshan South, and Hongqiao (Shanghai). A specific section is the railway part between two consecutive stations; for example, Nanjing-Zhenjiang Section is the railway part between Nanjing South Station and Zhenjiang West Station that is regarded as the 1st section on the Hu-Ning Part. The lengths of the six sections starting from Nanjing-Zhenjiang Section are 65110 m, 61050 m, 56400 m, 26810 m, 31350 m, and 43570 m, respectively. Its daily service starts at 6:30 am and ends at 11:30 pm. The planned timetable is obtained from Beijing-Shanghai High-speed Line Timetable Planning Lab, Beijing Jiao Tong University. As currently planned, there are 60 trains in the initial timetable. The trains from Beijing-Shanghai line are called self-line trains, while those from Riverside line are called cross-line trains. Fourteen high-speed trains from Beijing-Shanghai line take  $L$  and  $DJ$  as the train number prefix, 38 quasi-high-speed trains from Beijing-Shanghai line take  $G$  as the train number, and 8 quasi-high-speed trains from Riverside line take  $K$  as the train number. As in reality, self-line trains have higher priority than cross-line trains. As for self-line trains, the high-speed trains have higher priority than the quasi-high speed trains. The costs per time unit delay for the three kinds of trains are set to 5, 3, and 1, respectively. The costs of three kinds of rerouting strategies for trains (canceling train, detouring train, and merging train) are set based on the train's priority and its delay time. Delay tolerance for trains is set to 30 min. The minimum separation time on track possession in each station is set to 1 min.

In Figure 5, the high-speed line, inter-city line, and existing line are represented by  $\overline{NH}$ ,  $\overline{NH}_{ic}$ , and  $\overline{NH}_e$ , respectively. Correspondingly, the capacities are  $C$ ,  $C_{ic}$ , and  $C_e$ , and the demands are  $D$ ,  $D_{ic}$ , and  $D_e$ . The unit of both capacity and demand is the number of trains. Suppose there is a speed restriction on  $\overline{NH}$ , resulting in that the capacity of  $\overline{NH}$  reduces to  $C'$ .  $\overline{NH}_{ic}$  and  $\overline{NH}_e$  are in normal service, with known residual capacity  $\Delta C_{ic} = C_{ic} - D_{ie}$  and  $\Delta C_e = C_e - D_e$ , respectively. Before rescheduling the timetable in case of speed restriction,  $C'$  needs to be evaluated.

Although there are two different train speeds on the Beijing-Shanghai Part in the planned timetable, which are high-speed (350 km/hr) and quasi-high speed (300 km/hr), the speed of trains will become uniform when the speed restriction is lower than 300 km/h. Thus, in the two deduction factors mentioned at the end of Line Capacity Evaluation section, only the stopover of trains needs to be considered, but not the meeting of the trains. The deduction factor for single high-speed train with stopovers with regard to high-speed train without stopovers can be obtained by the following equation [24]:

$$\epsilon^{\text{high,stop}} = \frac{I + T}{I}, \quad (23)$$

where  $I$  denotes the minimum headway between consecutive trains and  $T$  denotes the minimum stopping time at

stations. Then, the average deduction factor for multiple high-speed trains can be obtained by the equation below [24]:

$$\epsilon_{\text{high}} = \alpha^{\text{high}} + \alpha^{\text{high,stop}} \epsilon^{\text{high,stop}}, \quad (24)$$

where  $\alpha^{\text{high}}$  denotes the ratio of high-speed trains without stopovers and  $\alpha^{\text{high,stop}}$  denotes the ratio of high-speed trains with stopovers. Given equations (1), (23), and (24) the capacity of high-speed rail line with trains of uniformed speed with stopovers can be obtained by the following equation:

$$C' = \frac{n_{\text{max}}}{\epsilon_{\text{high}}} = \frac{1440 - t_{\text{maintenance}} - t_{\text{inefficacy}}}{\epsilon_{\text{high}} I}, \quad (25)$$

where the headway  $I$  is calculated by (Number of blocks \* Length of block + Length of Train)/Train's average speed [24].

In the case study,  $T$  is 3 minutes. Since the capacity of the sections with the most stopovers restrains the capacity of the whole line,  $\alpha^{\text{high}} = 16/43$ ,  $\alpha^{\text{high,stop}} = 27/43$  are calculated in these sections. As planned,  $t_{\text{maintenance}} = 6$  hours (0:00 AM to 6:00 AM).  $t_{\text{inefficacy}} \approx \text{dis}/v$  is related to the total distance  $\text{dis}$  (284 km in this case) and the restricted train speed  $v$ . The lower the restricted speed is, the longer the inefficacy will be. The capacity of the Hu-Ning Part in various speed restrictions is calculated by (23)–(25), and the results are shown in Table 3.

In the case study, demand  $D = 43$ . It can be seen from Table 3 that  $C'$  decreases as the speed restriction gets lower. But it never gets smaller than  $D$ , which means the demand of Hu-Ning Part can always be fulfilled even when the speed is restricted down to 50 km/hr. The reason why the demand is so small may be that the obtained timetable is designed for the first step operation of the Beijing-Shanghai line, at which time of the high-speed network in blueprint has not formed yet.

Since  $C' \geq D$  holds in all the speed restriction cases, rerouting actions will not be taken at the first place as stated in the Rerouting Management section. The arrival/departure time adjustment optimization is executed, and the results are given in Table 4, where  $\eta_1$  denotes the number of trains with a delay not less than 30 mins,  $\eta_2$  denotes the number of trains with a delay not less than 60 mins,  $\lambda$  denotes the number of trains entering the last section beyond service hour, and  $\gamma$  denotes the number of delayed self-line trains. The values in brackets are the results with no rescheduling.

It can be seen from Table 4 that no train will be delayed for more than 30 mins when the speed restriction takes place in 1st section or 1st and 2nd sections, and the speed limitation is higher than 100 km/hr. There is a finding that the delay of 1st and 2nd sections speed restriction is more serious than that of 1st section speed restriction in terms of all the metrics. The results suggest that the longer the restricted distance is or the lower the speed limitation is, the more serious the delay will be. In addition, it can be seen from the results that both the total delay time and the number of delayed trains obtained with the proposed approach are less than those obtained without rescheduling.

As proposed in the Decision Tree-based Result Assessment section, a decision tree is used to evaluate the rescheduled timetable. In the case study, it is assumed that the

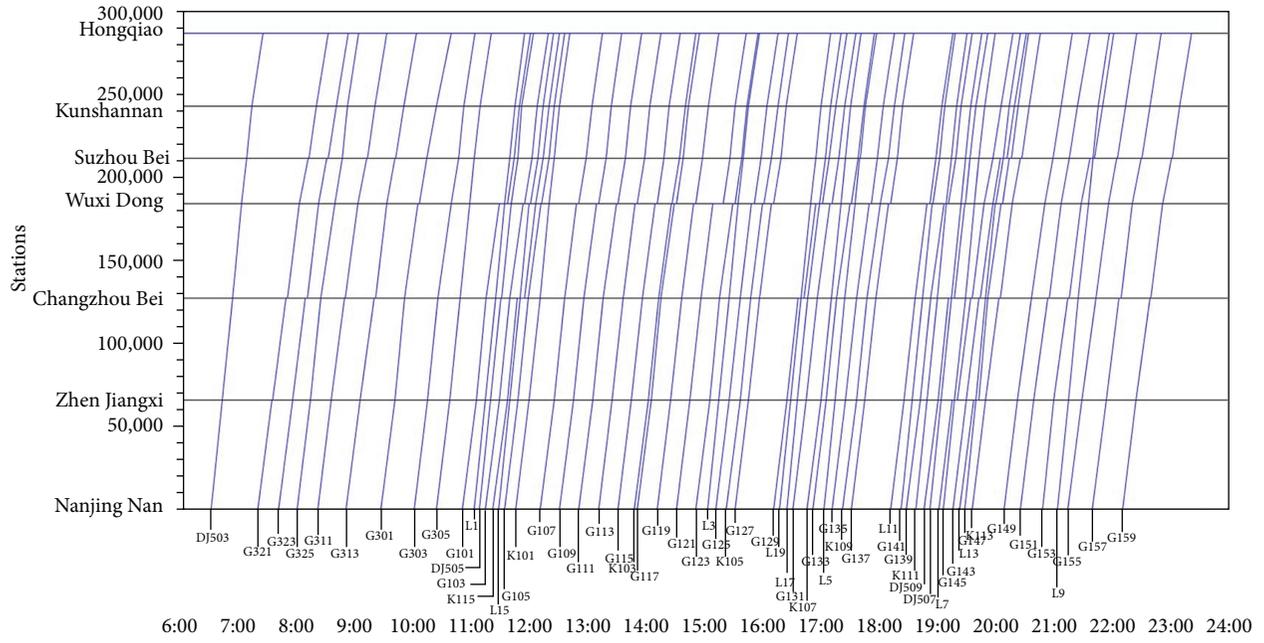


FIGURE 4: The planned timetable of Hu-Ning Part.

TABLE 3: Capacity of Hu-Ning part in speed restrictions.

Restricted speed (km/hr)	$t_{\text{inefficacy}}$ (hr)	$I$ (min)	$n_{\text{max}}$	$\epsilon_{\text{high,stop}}$	$\epsilon_{\text{high}}$	$C'$
300	N/A	2.3	469	2.3	1.82	257
200	1	2	510	2.5	1.94	262
100	2	2.2	436	2.36	1.85	235
80	3	2.7	333	2.1	1.69	197
50	6	2.4	300	2.25	1.78	168

TABLE 4: Evaluation results on the rescheduled timetable.

Restricted section(s)	Speed limitation (km/h)	$\eta_1$	$\eta_2$	$\lambda$	$\gamma$	Total delay (min)	Calculation time(s)
1st section	300	0 (0)	0 (0)	0 (0)	1 (14)	3.81 (23.1)	4.83
	200	0 (0)	0 (0)	0 (0)	25 (52)	98.21 (259.6)	5.24
	150	0 (0)	0 (0)	0 (0)	47 (52)	384.81 (810.2)	18.69
	100	0 (0)	0 (0)	0 (0)	52 (52)	1158.81 (1605.8)	9.22
	80	23 (60)	0 (0)	0 (0)	52 (52)	1743.31 (2017.6)	6.63
	<b>50</b>	<b>40 (0)</b>	<b>20 (60)</b>	<b>1(1)</b>	<b>52 (52)</b>	<b>3501.31 (3775.23)</b>	<b>6.69</b>
1st and 2nd sections	300	0 (0)	0 (0)	0 (0)	4 (52)	7 (41.44)	5.22
	200	0 (0)	0 (0)	0 (0)	47 (52)	383.13 (508.28)	8.13
	150	0 (0)	0 (0)	0 (0)	52 (52)	1125.28 (1264.98)	6.19
	100	60 (60)	0 (0)	0 (1)	52 (52)	2640.14 (2779.05)	5.28
	<b>80</b>	<b>18 (0)</b>	<b>42 (60)</b>	<b>1(1)</b>	<b>52 (52)</b>	<b>3776.15 (3911.99)</b>	<b>5.81</b>
	<b>50</b>	<b>0 (0)</b>	<b>60 (60)</b>	<b>3(3)</b>	<b>52 (52)</b>	<b>7180.28 (7320.87)</b>	<b>5.67</b>
All (1st–6th) sections	300	0 (0)	0 (0)	0 (0)	15 (52)	54.05 (93.63)	5.13
	200	11 (46)	0 (0)	0 (0)	52 (52)	1359.95 (1435.96)	3.7
	150	51 (0)	9 (60)	0 (1)	52 (52)	3065.95 (3141.71)	3.72
	<b>100</b>	<b>0 (0)</b>	<b>60 (60)</b>	<b>2(3)</b>	<b>52 (52)</b>	<b>6476.95 (6553.22)</b>	<b>4.27</b>
	<b>80</b>	<b>0 (0)</b>	<b>60 (60)</b>	<b>4(5)</b>	<b>52 (52)</b>	<b>9035.95 (9111.78)</b>	<b>3.67</b>
	<b>50</b>	<b>0 (0)</b>	<b>60 (60)</b>	<b>11 (15)</b>	<b>52 (52)</b>	<b>16716.15 (16787.6)</b>	<b>4.14</b>

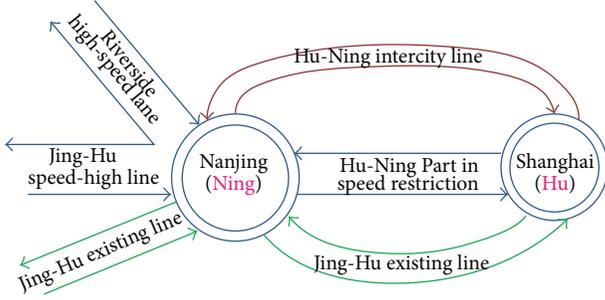


FIGURE 5: The railway network around Nanjing (Ning).

decision tree only comprises of one node, which is  $\lambda = 0$ . In other words, only if all the trains can enter the last section in service hour, the rescheduled result can be accepted. With this single-node decision tree, the unaccepted results are identified in Table 4, including 1st section-50 km/hr, 1st and 2nd sections-80 km/hr, 1st and 2nd sections-50 km/hr, all sections-100 km/hr, all sections-80 km/hr, and all sections-50 km/hr. The results imply that  $C'$  is not fully utilized.

As proposed in DSM, if a result is not accepted with result assessment, it goes back to the rerouting management module. In the case study, two rerouting strategies are considered, for  $C' \geq D$ . One is to merge/detour/cancel some trains, represented by RS-1. The other one is to advance departure time, represented by RS-2.

- (i) RS-1: In the optimization model of the rerouting management layer, we assume residual capacities  $\Delta C_{ic} = 3$  and  $\Delta C_e = 2$ . Advanced departure time is not allowed. Hence, the calculated maximum capacity is not applicable with this strategy. The strategy RS-1 is tested with the minimum rerouted number of trains  $\delta$  increasing from 1 to 13.
- (ii) RS-2: The departure time is allowed to be advanced by relaxing the second constraint of interval restrictions in the time adjustment-layer optimization model.

The comparison results are shown in Table 5.

It can be seen from Table 5 that detouring trains strategy has a high priority, merging trains strategy has a medium priority, and canceling trains strategy has a low priority according to the costs, respectively. The K-trains (K101, K103, K105, K107, K109, K111, K113, and K115) and G-trains (G149, G151, G153, G155, G157, and G159) take a detour or are cancelled because that the K-trains are cross-line trains that have lower  $c^{\text{detour}}$  and  $c^{\text{cancel}}$  than other trains, and the G-trains that depart from Nanjing South after 8 pm have a lower  $c^{\text{detour}}$  and  $c^{\text{cancel}}$  than other trains. The detoured train G151 is replaced by L9 when the speed limitation is lower than 80 km/h, and the speed restriction takes place in all sections. This may be explained that  $c^{\text{detour}}$  and  $c^{\text{cancel}}$  are related not only with the train priority but also with the speed limitation, and L9 departs after G151 from Nanjing South.

In addition, there is a finding that the results of RS-2 outperform RS-1 in terms of all the metrics. Although the results obtained by using RS-1 are better than those before

rerouting management shown in Table 4, the improvement is not big. The difference may be only resulted from the reduction of trains. Consequently, half of the results are still not acceptable based on the decision tree. In contrast, using RS-2 results is significant improvement for all the scenarios. Not only all the rescheduled timetables are acceptable, the values of all the metrics lower down considerably. The comparison result implies that to increase the utilization of line capacity is more helpful than to reduce demand when the train departure is sparse.

## 6. Conclusion and Future Work

This paper proposed a DSM for dispatchers to deal with timetable rescheduling in case of extensive speed restriction. The main contributions include the following. (1) The rerouting optimization is incorporated into the timetable rescheduling with a bilayer optimization model; (2) the line capacity calculation is innovatively used in solving timetable rescheduling problem to guide the selection of rerouting actions and parameters; (3) dispatchers are effectively involved into the iterative timetable rescheduling so that their valuable experience can be leveraged, while most research work on timetable rescheduling focuses on the optimization model itself.

The proposed decomposition into two optimization layers, including rerouting management and arrival/departure time adjustment, makes the timetable rescheduling optimization problem solvable even in a complex setting with an acceptable computation time. In addition, interposition and guidance of dispatchers are introduced into the optimization procedure with the decomposition, resulting in a more practical approach for real-world application. Despite these advantages, the optimality of the original timetable rescheduling optimization problem may be sacrificed in a sense due to the decomposition. The timetable rescheduling usually involves multiple objectives, of which the optimal solutions constitute a Pareto-front. When making the decision on which trains to be rerouted in the rerouting management layer, there is no way of considering the potential delay that each train will induce into the timetable. So it is possible that a good solution is lost with rerouting an improper train, thereby losing the corresponding Pareto-optimal point. Fortunately, the arrival/departure time adjustment layer guarantees to find the global optimal solution. Hence, a Pareto-optimal solution will be found in the end. In summary, the optimality is sacrificed in the sense that some Pareto-optimal solution is missed. This is acceptable because it is hard to say that one objective is more important than another in daily operation.

The proposed mechanism, methods, and models are simulated on the busiest part of the Beijing-Shanghai high-speed rail line. The proposed DSM showed good performance in the sense that the costs of delay and rerouting can be balanced and optimized iteratively. Additionally, a significant finding from the case study is that the most important step is to increase the utilization of the rail line in case of extensive speed restriction.

TABLE 5: Rescheduling results comparison on different rerouting strategies.

Rerouting strategy	Restricted section(s)	Speed limitation (km/h)	Merged trains	Detoured trains	Cancelled trains	$\eta_1$	$\eta_2$	$\lambda$	$\gamma$	Total delay (m)	Calculation time(s)
RS-1											
$\delta = 1$	1st section	50		G157, G159, K113, K105, K103		35	20	0	50	3217.6	8.25
	1st and 2nd sections	80		G153, G155, G157, G159, K113		14	41	0	48	3484.91	6.77
$\delta = 5$	All (1st-6th) sections	50		G153, G155, G157, G159, K113		0	55	0	48	6608.88	6.75
		100		G151, G153, G155, G157, G159		0	55	0	47	5971.05	4.89
$\delta = 6$	All (1st-6th) sections	80		G153, L9, G155, G157, G159		0	55	0	47	8300.7	4.88
		50		<b>G153, L9, G155, G157, G159</b>		<b>0</b>	<b>55</b>	<b>6</b>	<b>47</b>	<b>15337.03</b>	<b>4.94</b>
$\delta = 7$	1st section	50	G141-DJ507	G157, G159, K113, K105, K103		34	20	0	49	3160.98	6.98
	1st and 2nd sections	80	G141-DJ507	G153, G155, G157, G159, K113		14	41	0	47	3424.73	5.61
$\delta = 8$	All (1st-6th) sections	50	G141-DJ507	G153, G155, G157, G159, K113		0	54	0	47	6491.93	5.81
		100	G141-DJ507	G151, G153, G155, G157, G159		0	54	0	46	5866.53	3.88
$\delta = 9$	All (1st-6th) sections	80	G141-DJ507	G153, L9, G155, G157, G159		0	54	0	46	8153.53	4.56
		50	<b>G141-DJ507</b>	<b>G153, L9, G155, G157, G159</b>		<b>0</b>	<b>54</b>	<b>6</b>	<b>46</b>	<b>15061.93</b>	<b>3.86</b>
$\delta = 10$	1st section	50	DJ505-G103, G141-DJ507	G157, G159, K113, K115, K101		35	18	0	48	3098.01	6.66
	1st and 2nd sections	80	DJ505-G103, G141-DJ507	G153, G155, G157, G159, K113		13	40	0	46	3362.55	5.55
$\delta = 11$	All (1st-6th) sections	50	DJ505-G103, G141-DJ507	G153, G155, G157, G159, K113		0	53	0	46	6366.15	5.27
		100	DJ505-G103, G141-DJ507	G151, G153, G155, G157, G159		0	53	0	45	5750.48	3.62
$\delta = 12$	All (1st-6th) sections	80	DJ505-G103, G141-DJ507	G153, L9, G155, G157, G159		0	53	0	45	7994.83	3.66
		50	<b>G141-DJ507</b>	<b>G151, G153, L9, G155, G157</b>	<b>G159</b>	<b>0</b>	<b>53</b>	<b>5</b>	<b>45</b>	<b>14789.83</b>	<b>3.55</b>
$\delta = 13$	1st section	50	DJ505-G103, G141-DJ507, G145-G143	G157, G159, K109, K115, K101		35	17	0	47	3037.26	6.77
	1st and 2nd sections	80	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K113		12	40	0	45	3302.73	5.27
$\delta = 14$	All (1st-6th) sections	50	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K113		0	52	0	45	6250.73	5.33
		100	DJ505-G103, G141-DJ507	G151, G153, G155, G157, G159	K113	0	52	0	45	5648.96	3.65
$\delta = 15$	All (1st-6th) sections	80	DJ505-G103, G141-DJ507	G153, L9, G155, G157, G159	G151	0	52	0	44	7850.66	4.03
		50	<b>DJ505-G103, G141-DJ507</b>	<b>G151, G153, L9, G155, G157</b>	<b>G159</b>	<b>0</b>	<b>52</b>	<b>5</b>	<b>44</b>	<b>14503.2</b>	<b>3.55</b>
$\delta = 16$	1st section	50	DJ505-G103, G141-DJ507, G145-G143, G129-G131	G157, G159, K101, K115, K103		34	17	0	46	2983.88	6.05
	1st and 2nd sections	80	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K103	K111	13	38	0	45	3227.95	4.36
$\delta = 17$	All (1st-6th) sections	50	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K103	K105	0	51	0	45	6128.98	4.75
		100	DJ505-G103, G141-DJ507, G145-G143	G151, G153, G155, G157, G159	K113	0	51	0	44	5552.45	3.67
$\delta = 18$	All (1st-6th) sections	80	DJ505-G103, G141-DJ507	G153, L9, G155, G157, G159	G151, K113	0	51	0	44	7706.5	3.67
		50	<b>DJ505-G103, G141-DJ507</b>	<b>G151, L9, G155, G157, G159</b>	<b>G149, G153</b>	<b>0</b>	<b>51</b>	<b>4</b>	<b>43</b>	<b>14235.3</b>	<b>3.52</b>

TABLE 5: Continued.

Rerouting strategy	Restricted section(s)	Speed limitation (km/h)	Merged trains	Detoured trains	Cancelled trains	$\eta_1$	$\eta_2$	$\lambda$	$\gamma$	Total delay (m)	Calculation time(s)
$\delta = 10$	1st section	50	DJ505-G103, G141-DJ507, G145-G143, G129-G131	G157, G159, K101, K115, K103	K113	33	17	0	46	2923.53	4.75
	1st and 2nd sections	80	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K103	K111, K109	13	37	0	45	3164.76	4.14
	All (1st–6th) sections	50	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K103	K105, K111	0	50	0	45	5996.68	4.39
		100	DJ505-G103, G141-DJ507, G145-G143	G151, L9, G155, G157, G159	G153, K113	0	50	0	43	5434.83	2.84
		80	DJ505-G103, G141-DJ507	G151, G159, L9, G155, G157	G153, G149, K113	0	50	0	43	7562.33	2.89
	50	<b>DJ505-G103, G141-DJ507</b>	<b>G149, G151, G153, L9, G157</b>	<b>G155, G159, K113</b>	<b>0</b>	<b>50</b>	<b>3</b>	<b>43</b>	<b>13959.0</b>	<b>2.81</b>	
$\delta = 11$	1st section	50	DJ505-G103, G141-DJ507, G145-G143, G129-G131	G157, G159, K105, K115, K103	K113, K111	32	17	0	46	2866.5	4.42
	1st and 2nd sections	80	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K115	K111, K109, K113	12	37	0	45	3106.78	3.83
		50	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K101	K105, K111, K113	0	49	0	45	5876.9	3.94
	All (1st–6th) sections	100	DJ505-G103, G141-DJ507, G145-G143	G151, L9, G155, G157, G159	G153, K113, K115	0	49	0	43	5332.31	3.25
		80	DJ505-G103, G141-DJ507, G145-G143	G149, G155, G159, L9, G157	K113, G153, G151	0	49	0	42	7423.16	3.39
	50	<b>DJ505-G103, G141-DJ507, G145-G143</b>	<b>G151, G153, L9, G155, G157</b>	<b>G159, G149, K113</b>	<b>0</b>	<b>49</b>	<b>2</b>	<b>42</b>	<b>13691.9</b>	<b>3.24</b>	
$\delta = 12$	1st section	50	DJ505-G103, G141-DJ507, G145-G143, G129-G131	G157, G159, K105, K115, K103	K113, K111, K109	31	17	0	46	2807.6	4.19
	1st and 2nd sections	80	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K101	K111, K109, K113, K115	12	36	0	45	3036.18	3.86
		50	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K103	K105, K109, K111, K113	0	48	0	45	5763.36	3.64
	All (1st–6th) sections	100	DJ505-G103, G141-DJ507, G145-G143	G151, L9, G155, G157, G159	G153, K113, K115, K109	0	48	0	43	5224.8	2.88
		80	DJ505-G103, G141-DJ507, G145-G143	G149, G155, G159, L9, G157	K101, K113, G153, G151	0	48	0	42	7265.58	2.64
	50	<b>DJ505-G103, G141-DJ507, G145-G143</b>	<b>G149, G151, G153, L9, G157</b>	<b>K105, K113, G155, G159</b>	<b>0</b>	<b>48</b>	<b>2</b>	<b>42</b>	<b>13413.8</b>	<b>2.53</b>	

TABLE 5: Continued.

Rerouting strategy	Restricted section(s)	Speed limitation (km/h)	Merged trains	Detoured trains	Cancelled trains	$\eta_1$	$\eta_2$	$\lambda$	$\gamma$	Total delay (m)	Calculation time(s)
$\delta = 13$	1st section	50	DJ505-G103, G141-DJ507, G145-G143, G129-G131	G157, G159, K105, K115, K103	K113, K111, K109, K107	31	16	0	46	2741.58	3.77
	1st and 2nd sections	80	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K103	K111, K109, K113, K115, K101	12	35	0	45	2972.0	3.41
		50	DJ505-G103, G141-DJ507, G145-G143	G153, G155, G157, G159, K103	K105, K109, K111, K113, K115	0	47	0	45	5647.45	3.46
	All (1st–6th) sections	100	DJ505-G103, G141-DJ507, G145-G143	G151, G153, L9, G155, G157	K115, K109, K113, K101, G159	0	47	0	43	5109.86	2.39
RS-2		80	DJ505-G103, G141-DJ507, G145-G143	G149, G151, L9, G157, G159	K115, K101, K113, G153, G155	0	47	0	42	7120.41	3.19
		50	<b>DJ505-G103, G141-DJ507, G145-G143</b>	<b>G151, G153, L9, G155, G157</b>	<b>K115, K105, K113, G149, G159</b>	<b>0</b>	<b>39</b>	<b>2</b>	<b>39</b>	<b>10841.9</b>	<b>2.95</b>
	1st section	50	—	—	—	0	1	0	10	118.63	65.47
	1st and 2nd sections	50	—	—	—	0	1	0	11	174.03	18.73
RS-2	All (1st–6th) sections	100	—	—	—	2	2	0	11	352.73	67.53
		80	—	—	—	3	1	0	52	806.1	13.22
		50	—	—	—	0	5	0	47	963.73	18.08
			—	—	—	0	12	0	49	2512.74	563.59

There remain some interesting areas to explore around the proposed DSM. Firstly, the speed restriction is set per day per section; how to make the setting more flexible is an issue. Secondly, the rolling-stock rebalancing is yet to be considered together with the timetable rescheduling. Moreover, the proposed DSM is designed for decision support purpose. Our ultimate goal is to design and develop a real-time decision support framework in the future.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work has been supported by National Natural Science Foundation of China (61074151), National Key Technology Research and Development Program (2009BAG12A10), and National High Technology Research and Development Program of China (2012AA112001).

## References

- [1] J. Jacobs, "Reducing delays by means of computer-aided "on-the-spot" rescheduling," *Computers in Railways IX*, WIT, Southampton, UK, vol. 15, pp. 603–612, 2004.
- [2] T. Norio, T. Yoshiaki, T. Noriyuki, H. Chikara, and M. Kunimitsu, "Train rescheduling algorithm which minimizes passengers' dissatisfaction," *Transactions of Information Processing Society of Japan*, vol. 46, no. 2, pp. 26–38, 2005.
- [3] A. D'Ariano, D. Pacciarelli, and M. Pranzo, "A branch and bound algorithm for scheduling trains in a railway network," *European Journal of Operational Research*, vol. 183, no. 2, pp. 643–657, 2007.
- [4] A. D'Ariano, F. Corman, D. Pacciarelli, and M. Pranzo, "Reordering and local rerouting strategies to manage train traffic in real time," *Transportation Science*, vol. 42, no. 4, pp. 405–419, 2008.
- [5] A. D'Ariano and T. Albrecht, "Running time re-optimization during real-time timetable perturbations," in *Computers in Railways X*, J. Allan, C. A. Brebbia, A. F. Rumsey, G. Sciotto, S. Sone, and C. J. Goodman, Eds., pp. 531–540, WIT, Southampton, UK, July 2006.
- [6] A. D'Ariano, M. Pranzo, and I. A. Hansen, "Conflict resolution and train speed coordination for solving real-time timetable perturbations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 208–222, 2007.
- [7] A. D'Ariano, F. Corman, D. Pacciarelli, and M. Pranzo, "Reordering and local rerouting strategies to manage train traffic in real time," *Transportation Science*, vol. 42, no. 4, pp. 405–419, 2008.
- [8] A. D'Ariano and M. Pranzo, "An advanced real-time train dispatching system for minimizing the propagation of delays in a dispatching area under severe disturbances," *Networks and Spatial Economics*, vol. 9, no. 1, pp. 63–84, 2009.
- [9] L. Wang, L.-M. Jia, Y. Qin, J. Xu, and W.-T. Mo, "A two-layer optimization model for high-speed railway line planning," *Journal of Zhejiang University: SCIENCE A*, vol. 12, no. 12, pp. 902–912, 2011.
- [10] S. Wegele and E. Schnieder, "Dispatching of train operations using genetic algorithms," in *Proceedings of the 9th International Conference on Computer-Aided Scheduling of Public Transport (CASPT '04)*, San Diego, Calif, USA, 2004.
- [11] J. Törnquist and J. A. Persson, "Train traffic deviation handling using tabu search and simulated annealing," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, p. 73, January 2005.
- [12] J. Törnquist and J. A. Persson, "N-tracked railway traffic re-scheduling during disturbances," *Transportation Research Part B*, vol. 41, no. 3, pp. 342–362, 2007.
- [13] I. Şahin, "Railway traffic control and train scheduling based on inter-train conflict management," *Transportation Research Part B*, vol. 33, no. 7, pp. 511–534, 1999.
- [14] L. Wang, Y. Qin, J. Xu, and L. Jia, "A fuzzy optimization model for high-speed railway timetable rescheduling," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 827073, 22 pages, 2012.
- [15] B. Adenso-Díaz, M. Oliva González, and P. González-Torre, "On-line timetable re-scheduling in regional train services," *Transportation Research Part B*, vol. 33, no. 6, pp. 387–398, 1999.
- [16] A. Higgins, E. Kozan, and L. Ferreira, "Optimal scheduling of trains on a single line track," *Transportation Research Part B*, vol. 30, no. 2, pp. 147–158, 1996.
- [17] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: a fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.
- [18] H. Guo, W. Wang, W. Guo, X. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [19] E. R. Kraft, "Jam capacity of single track rail lines," *Proceedings of the Transportation Research Forum*, vol. 23, no. 1, pp. 461–471, 1982.
- [20] B. Szpigel, "Optimal train scheduling on a single track railway," *Operational Research*, vol. 72, pp. 343–352, 1972.
- [21] D. Jovanović and P. T. Harker, "Tactical scheduling of rail operations. The SCAN I system," *Transportation Science*, vol. 25, no. 1, pp. 46–64, 1991.
- [22] N. Welch and J. Gussow, "Expansion of Canadian National Railway's line capacity," *Interfaces*, vol. 16, no. 1, pp. 51–64, 1986.
- [23] UIC leaflet 406, *Capacity*, UIC, Paris, France, 2004.
- [24] H. Yang, *Theory of Railway Transportation Planning*, China Railway Publishing House, 1999.

## Research Article

# An Operation Optimization for Express Freight Trains Based on Shipper Demands

Yuzhao Zhang<sup>1,2</sup> and Yusong Yan<sup>2</sup>

<sup>1</sup> School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

<sup>2</sup> School of Transportation & Logistics, Southwest Jiaotong University, Chengdu 610031, China

Correspondence should be addressed to Yuzhao Zhang; yuzhaozhang@126.com

Received 18 November 2013; Revised 16 December 2013; Accepted 24 December 2013; Published 12 January 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Y. Zhang and Y. Yan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The supreme goal of railway express freight transportation is to meet the various demands of shippers. This paper considers the optimization problem of the operation plan for express freight trains based on shipper demands, including transport volume, service frequency, and transit time. An integer programming model that takes maximum shipper demands as optimized objective is proposed to realize the assignment of trains to candidate paths. In this model, the number of marshalling cars for express freight train is set as a decision variable rather than a constant to better satisfy the demand characteristics. A heuristic solution algorithm based on genetic algorithm is then designed. Finally, a numerical case is applied to verify the feasibility of the model and the algorithm.

## 1. Introduction

Express freight has an important role in the whole freight transportation system. With the constant increase in energy cost, the increasing seriousness of traffic congestion, and environmental pollution, the express freight transportation system that takes aviation and expressway as the main modes reflects many problems [1]. Many countries have paid more attention to the development of rail express freight transportation, which is more energy-saving and environment friendly [2]. Being the foundation of the railway freight transportation organization, operation plan of express freight trains is to determine the set of trains to satisfy the express freight demands. Compared with bulk freight, express freight has the characteristics of small batch, excellent timeliness, and high value, and the shippers have higher requirements on service quality, but lower sensitivity on the transport cost. Consequently, the characteristics and shipper diversified demands of express freight should be considered when determining and optimizing the train operation plan.

In the past decades, various attempts have been made to gain optimal transportation scheme for railway express freight demands. These attempts mainly involve three kinds:

design of formation plan for direct freight trains, service network design of express freight, and determination of operation plan for express freight trains.

Formation plan of direct freight trains determines the transport scheme of car flows by direct trains, including direct trains in the loading place and the technical station [3]. Many express freight trains are direct trains, and freight demands can be converted into car flows; thus, its determination method can provide references for the optimization of operation plan for express trains. Lin et al. [3] have researched on train formation plan problem for a long time and achieved mature modeling approaches that take minimum time consumption of car flows as objective and 0-1 programming model as basis. Crainic et al. [4] established a mixed integer planning model for railway freight transportation to realize the strategy plan of enterprise resources, which aims to minimize the operating cost of transportation enterprise. Marín and Salmerón [5] formulated a model of assigning the trains and the demand of cars to routes for rail freight networks and designed different heuristic algorithms involved in simulated annealing, Tabu search, and descending to respond to large scale network. Yaghini et al. [6] proposed a population-based genetic algorithm for railway

blocking problem and then compared the solution quality and the computational time of the algorithm with the optimization solver software, which showed that the proposed algorithm has high efficiency and effectiveness. Martinelli and Teng [7] built a nonlinear integer programming model for railway operation plan taking the minimum operation time as objective and presented neural network to solve the model.

Designing reasonable service network to meet express freight demands is another important aspect. Crainic [8] classified the service network design of freight transportation enterprise into service frequency design and dynamical service network design and established two kinds of frequency design models with dynamic service network design model based on space-time network. Lin and Chen [9] contrived hub-and-spoke service network with the limitation of transit time for express delivery enterprises. Although this method has no direct influence on railway express transportation, the idea can provide experience for the service network design of railway express freight. Ceselli et al. [10] established three models for the express service network of Swiss federal railways applying different methods and designed corresponding solution methods that include commercial solver approach, branch-and-cut approach, and column generation-based approach.

In recent years, the number of studies on express freight train operation plan has been increasing. Ben-Tal et al. [11] proposed an operating-plan model (OPM) for the Santa Fe Railway and designed a combination algorithm using genetic and tabu searches; the application of the OPM reduced 4 percent to 6 percent of the operation costs for the company. Wang and Liu [12] discussed the operation optimization problem of special parcel trains without fixed demands and capacity constraint and determined the running sections, number, and marshalling of trains by establishing a linear programming model. Yano and Newman [13] took the minimization sum of transportation cost and holding cost as objective and formulated an optimization model and an algorithm for the scheduling problem of trains and containers with due dates. Crevier et al. [14] constructed a bilevel model for the operation plan of rail freight transportation comprehensively considering the pricing policy and network capacity and then designed an exact solution approach based on branch-and-bound algorithm. Guo et al. [15] put forward a linear integer programming model for the operation plan of scheduled transit trains and solved the problem using Lingo software.

These studies have provided useful references for the design and optimization of express freight trains operation plan. Nonetheless, most of the present researches formulate models based on the operation with less consideration on the shipper demands. The number of marshalling cars for express freight trains is usually set as a constant. According to the characteristics of express freight, the operation plans of express freight trains with flexible marshalling are more suitable. Furthermore, the problem of operation plan for express freight trains is often solved by the exact solution methods, which cannot work when the network scale is large. This paper focuses on shipper demands of express freight and proposes a model considering flexible marshalling of express

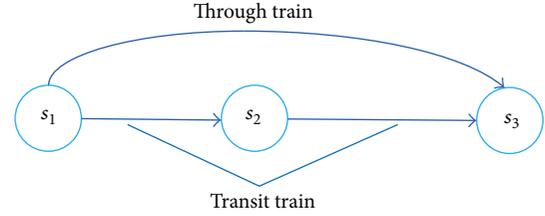


FIGURE 1: Illustration of express freight train kinds.

freight trains, and a heuristic algorithm will be designed to solve the model.

The remainder of the paper is organized as follows. An optimization model for operation plan of express freight trains based on shipper demands is established in Section 2, and then the solution algorithm on the basis of genetic algorithm is designed in Section 3. In Section 4, a numerical example is provided to verify the effectiveness of the model and the algorithm. The last section gives some conclusions and the possible areas of further research.

## 2. Model Formulation

**2.1. Problem Statement.** Express freight trains can be sorted into two classifications, namely, through trains and transit trains, depending on their operation. For example, Figure 1 shows a simple railway network composed of three organization stations of express freight and two sections. The trains with origin and destination stations ( $s_1$  and  $s_3$ ), which only consist of car flows from  $s_1$  to  $s_3$  and without reorganization in station  $s_2$ , belong to the first classification, whereas the others belong to the second classification. By organizing transit trains, the converging time of car flows can be reduced, but the transit time and transit cost will increase. This paper only considers the operation problem associated with the first classification express freight trains. In the light of shipper demands with consideration to the benefit of railway transportation enterprise, the origin and destination stations, running paths, number of marshalling cars, and running frequency of express freight trains will be determined under the network condition. In this paper, the number of marshalling cars was set as a decision variable, so the operation plan with flexible marshalling was formulated.

Aside from the basic requirement on security, shippers of express freight have higher expectation on transportation services, including satisfaction degree of demand, transport time, and transport convenience but have lower sensitivity to transport cost [15]. This paper mainly considers these expectations of shippers when establishing the model, which takes maximum transport volume as the objective and meeting the requirement of transport time and convenience as constraints. In Particular the transport convenience constraint is realized by making the running frequency of express freight trains more than the desired service frequency of shippers. In addition, constraints in the number of marshalling cars and station capacity, as well as section capacity are also considered.

The operation plan of express freight trains is a complex integrated optimization problem, and the following assumptions are made to reduce the difficulty in establishing the model.

- (1) The stations being considered in this paper are only the organizing stations and handling stations of express freight, which have favorable distribution conditions, and can meet the requirement of departure and arrival for the whole train.
- (2) The classification and transportation conditions of goods are not considered, while supposing the grades of all the express freight trains are the same.
- (3) The loading capacity of each car is identical, indicating the freight volume loaded on every car is the same. Furthermore, freight demands are converted into car flows of each day, and the traction tonnages of each section are also denoted by the traction cars.
- (4) The candidate paths of each express freight demand are determined, and the running path of the same demand is consistent.

**2.2. Definitions and Notations.** We define the following notations to describe the proposed model:

$G(S, E)$  : railway physical network;

$S$  : set of station in  $G$ , and we denote a station by  $s$ ;

$E$  : set of edge in  $G$ , and we denote an edge by  $e$ ;

$Q$  : set of express freight demand  $OD$ , and we denote a demand  $OD$  by  $q$ ;

$O_q$  : origin station set of  $q$ ;

$D_q$  : destination station set of  $q$ ;

$n^q$  : freight volume of  $q$ , and we transfer it into number of cars;

$f^q$  : desired service frequency of  $q$ ;

$t^q$  : desired transit time of  $q$ , including the overall time from origin to destination;

$P_q$  : set of path for  $q$ ; we denote a path by  $p$ , which can be expressed by a serial set of stations, and the first element is the original station of the path, while the final element is the destination station;

$P$  : set of all paths, and  $P = \cup_{q \in Q} P_q$ ;

$t_p$  : transit time by path  $p$ ;

$f_p$  : service frequency on path  $p$ , and it is expressed by the number of trains operating on path  $p$  every day, while its value is within a range of nonnegative integers;

$n(s)$  : departure and arrival capacity of station  $s$ ,  $\forall s \in S$ ;

$n(e)$  : carrying capacity of edge  $e$ ,  $\forall e \in E$ .

We define two decision variables as follows.

$x_p^q$  : binary variable that denotes the choice of path for demand  $q$ ;  $x_p^q$  equals to 1 if freight  $q$  was distributed to path  $p$ , indicating express freight trains are operated on path  $p$ , while 0 otherwise.

$y_p$  : number of marshalling cars on path  $p$ ,  $\forall p \in P$ , if  $x_p^q = 0$ ; then  $y_p = 0$ , otherwise if  $x_p^q = 1$ , then  $y_p \in N$ .

**2.3. Objective Function.** This paper takes maximum transport volume as the objective function, so that the operation plan can transport freights as much as possible to satisfy shipper demands. The objective function can be expressed as follows

$$\text{Max } Z = \sum_{q \in Q} \sum_{p \in P_q} x_p^q y_p f_p. \quad (1)$$

#### 2.4. Constraint Conditions

**2.4.1. Service Frequency Constraint.** The service frequency of express freight trains should not be less than the minimum desired frequency of shippers. The service frequency of demand  $q$  on path  $p$  can be expressed by  $\sum_{p \in P_q} x_p^q f_p$ ; thus, for  $\forall q \in Q$ , the constraint is expressed by the following inequality:

$$\sum_{p \in P_q} x_p^q f_p \geq f^q. \quad (2)$$

Only if path  $p$  was selected to transport freight, there will be trains operating on this path, and the service frequency  $f_p$  belongs to a range of positive integers which are bigger than  $f^q$ . Otherwise,  $f_p$  equals 0. Consequently, the following formula is obtained:

$$f_p = \begin{cases} 0, & \text{if } x_p^q = 0 \\ f_p \geq f^q, f_p \in N, & \text{if } x_p^q = 1. \end{cases} \quad (3)$$

**2.4.2. Transit Time Constraint.** Due to the high timeliness of express freight, the transit time should be within the requirement time range in the process of express freight transportation. The expression is as follows:

$$\sum_{p \in P_q} x_p^q t_p \leq t^q. \quad (4)$$

**2.4.3. Number of Marshalling Cars.** The number of marshalling cars for express trains will be equal to 0 if  $x_p^q = 0$ . Otherwise, if  $x_p^q = 1$ , the number will be within a range of positive integers, of which the upper limit value is the minimal traction cars of each section on the path and the lower limit value should ensure the benefit of railway enterprise. As a result, the number of marshalling cars for express freight trains can be expressed by the following formula:

$$y_p = \begin{cases} 0, & \text{if } x_p^q = 0 \\ m_p^{\min} \leq y_p \leq m_p^{\max}, y_p \in Z^+, & \text{if } x_p^q = 1, \end{cases} \quad (5)$$

Upper layer	1	0	0	1	1	0	1	0	0	1	0	1	Choice of path
Lower layer	20	0	0	25	22	0	30	0	0	24	0	20	Number of marshalling cars

FIGURE 2: Illustration of chromosome coding approach.

where  $m_p^{\min}$  and  $m_p^{\max}$ , respectively, denote the minimum number and the maximum number of marshalling cars for express freight trains on path  $p$ ,  $\forall p \in P$ .

**2.4.4. The Relationship between Transport Volume and Demand.** For  $\forall q \in Q$ , the overall volume transported by all paths should not exceed the amount of transportation demand, which can be expressed as follows:

$$\sum_{p \in P_q} x_p^q y_p f_p \leq n^q. \quad (6)$$

**2.4.5. Car Flow Constraint.** According to the fourth assumption, one group car flow can only be allocated to one path then, for  $\forall q \in Q$ , we can get the following inequality:

$$\sum_{p \in P} x_p^q \leq 1. \quad (7)$$

When  $\sum_{p \in P} x_p^q = 1$ , the car flow of  $q$  is distributed to a path  $p$ , indicating  $q$  is transported by trains operating on path  $p$ . If  $\sum_{p \in P} x_p^q < 1$ , then  $q$  is not distributed to any path, indicating it is not transported.

**2.4.6. Carrying Capacity Constraint of Each Section.** The number of running trains on each section should not exceed its carrying capacity. Set  $\delta_p^e$  as an auxiliary binary variable;  $\delta_p^e$  takes value 1 if  $e \in p$  and 0 otherwise; then, the constraint of carrying capacity can be expressed as follows:

$$\sum_{p \in P} x_p^q \delta_p^e f_p \leq n(e). \quad (8)$$

**2.4.7. Departure and Arrival Capacity Constraint.** The total number of trains departing from station  $s$  and arriving at it could not exceed its departure and arrival capacity, which can be expressed by the following formula:

$$\sum_{q \in Q} \sum_{p \in P_q} x_p^q \varphi_p^s f_p \leq n(s), \quad (9)$$

where for  $\forall s \in O_q \cup D_q$ , incidence vector  $\varphi_p^s$  takes value 1 if  $s \in p$  and 0 otherwise.

### 3. Algorithm Design

The proposed model is a nonlinear integer programming model, which belongs to NP problem. It is difficult to solve by traditional exact solving methods and commercial solver,

especially when the network scale is large, getting the solution is nearly impossible [16]. Thus, we design a heuristic genetic algorithm for the problem to get a satisfying solution within a tolerable time range.

**3.1. Coding Approach.** Aiming at the two-group decision variables of the model, a bilayer coding structure [17, 18] is adopted in this paper. A binary number embedded with the upper layer coding denotes the choices of paths, where the value 1 indicates the path being selected and 0 otherwise. A nonnegative integer embedded with the lower layer coding indicates the number of marshalling cars for express freight train; if a path is chosen, then the integer is within  $[m_p^{\min}, m_p^{\max}]$ ; otherwise, the integer is zero. The length of chromosome equals the total number of candidate paths  $M$ .

Figure 2 illustrates a bi-layer coding chromosome, the length of which is 12. The number range of marshalling cars is [20, 30], and six paths are chosen to operate express freight trains, whose numbers of marshalling cars are, respectively, 20, 25, 22, 30, 24, and 20.

**3.2. Initialization of Group.** Get the value of the upper layer gene by generating a binary array randomly with the length of  $M$  and then corresponding gene value of the lower layer is formatted in view of the binary array. Supposing the population size is  $N$ , the initial population can be obtained by the above mentioned way. All chromosomes have to be tested according to the constraints. If an individual is infeasible, necessary adjustment will be carried out. Particularly, the total number of chosen paths for any demand  $q$  should not exceed 1; in order to enhance the solving efficiency, we let the initial number equal 1. The  $x_p^q$  taking value 1 associated with demand  $q$  should be reduced by 1 with descending order if the total number of chosen paths exceeds 1, until the condition is satisfied. Contrarily, if the total number of chosen paths for demand  $q$  equals 0, then choose a candidate path randomly associated with  $q$ . The number of marshalling cars for each path should be adjusted with the change of  $x_p^q$ . As showed in Figure 3, the values enclosed in each couple curly braces represent the choice of each demand, and a new chromosome can be obtained after the adjustment of the original one if the values of  $m_p^{\min}$  and  $m_p^{\max}$  are 20 and 30, respectively. Moreover, chromosomes should also be tested and adjusted to ensure their feasibility after any genetic operation.

**3.3. Fitness Function.** For the proposed model, the objective function is a maximum problem, while the function value of which is nonnegative, so the objective function of the proposed model is taken as the fitness function.

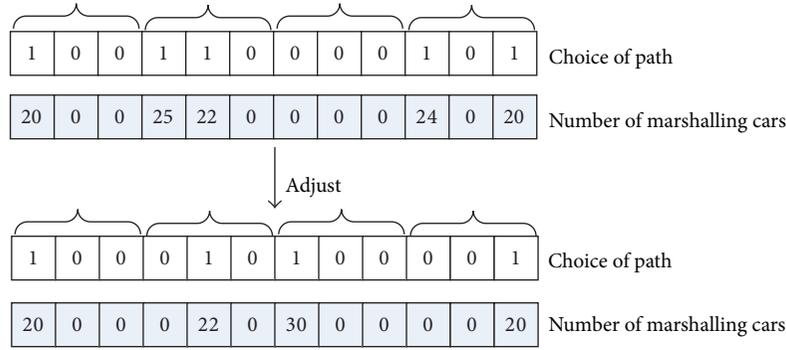


FIGURE 3: Adjustment of unfeasible chromosome.

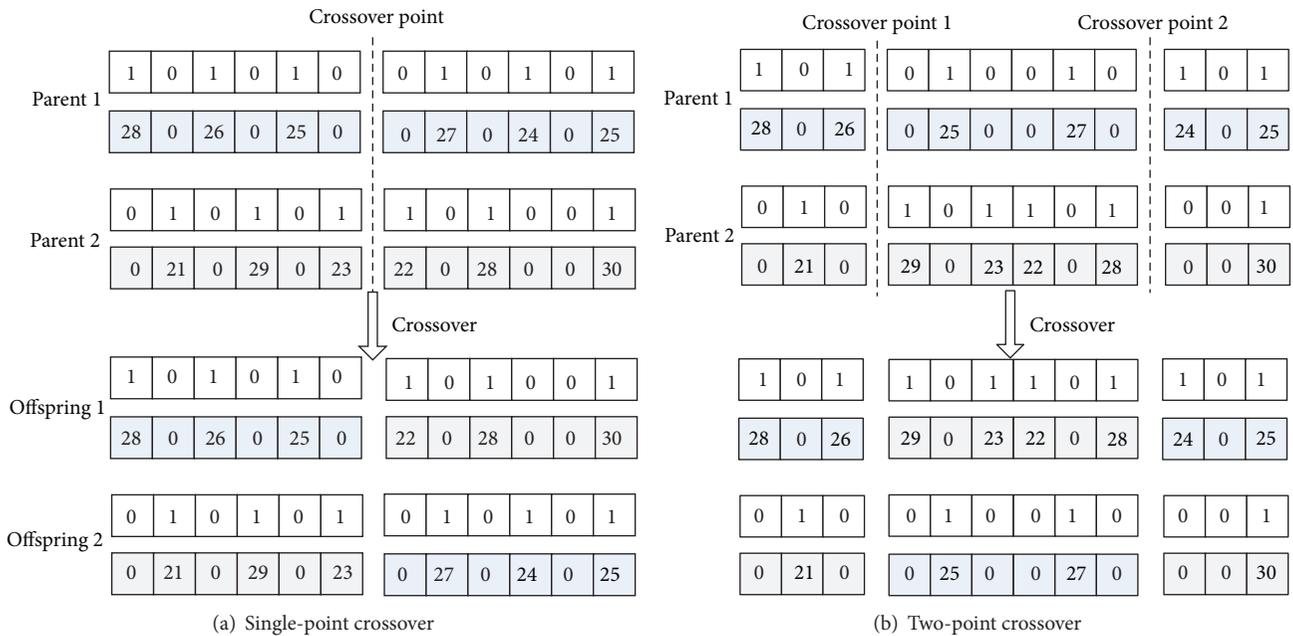


FIGURE 4: Illustration of crossover operation.

3.4. *Selection Operator.* Roulette selection method is adopted in this paper. In each round, a uniform random number between zero and one is generated as the selection finger to determine the candidate individuals. The best individual preservation strategy is also applied to ensure the present best individual not being destroyed by genetic operation.

3.5. *Crossover Operator.* In view of the bi-layer coding mode, a synchronous crossover approach of the upper layer gene string and the lower layer gene string is adopted to guarantee the feasibility of offsprings. Meanwhile, single-point and two-point crossover methods are comprehensively applied. The procedure of the crossover operation is as follows.

*Step 1.* Set the crossover probability  $P_c$  and generate a random number  $\lambda$  of uniform distribution within  $[0, 1]$  to be the judgment criterion.

*Step 2.* Do crossover operation if  $\lambda \leq P_c$ , where if  $\lambda \leq P_c/2$ ; then select a crossover point randomly and exchange the gene strings associated with both the upper layer and the lower layer beside the crossover point of parental chromosomes to get two offspring chromosomes. If  $(P_c/2) < \lambda \leq P_c$ , two crossover points are selected randomly, and the gene strings between the two crossover points of parental chromosomes are exchanged to obtain offspring chromosomes.

Figure 4 gives an illustration of single-point crossover operation and two-point crossover operation.

3.6. *Mutation Operator.* Due to the fact that the number of marshalling cars is influenced by the choice of the path, which means that the values of the lower layer gene string are influenced by the values of the upper layer gene string, the following algorithm is designed.

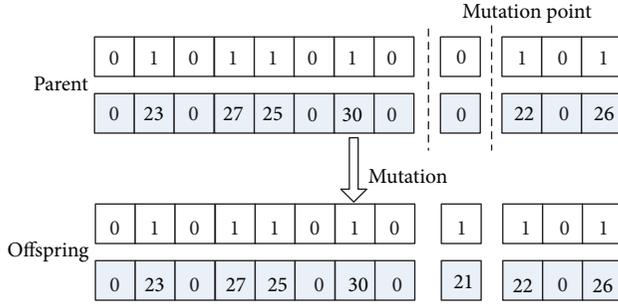


FIGURE 5: Illustration of mutation operation.

*Step 1.* Mutation probability  $P_m$  is set, and a random number  $\lambda'$  of uniform distribution at  $[0, 1]$  is generated to be the judgment criterion.

*Step 2.* One mutation point is selected randomly to be the mutation position of the upper layer gene and the lower layer gene.

*Step 3.* Do mutation operation if  $\lambda' \leq P_m$ . Firstly, mutation is operated on the upper layer chromosome, and then the value of the lower layer gene is adjusted according to the upper layer. If the value of the upper layer gene changes from 1 to 0, then the value of the corresponding lower layer gene turns from a positive integer to 0. Inversely, if the value of the upper layer gene changes from 0 to 1, then the value of the corresponding lower layer gene turns from zero to a positive integer.

The illustration of the mutation operation is shown in Figure 5.

#### 4. Numerical Example

The railway physical network shown as Figure 6 is structured to verify the proposed model and the algorithm. In this figure, the circles indicate the railway stations that deal with the services of express freight trains, while the edges indicate the sections. Relative parameter values are shown in Tables 1 and 2. The information of express freight demand  $OD$ , demand volume, desired minimum service frequency, and maximum transit time of shippers, as well as candidate paths of each demand is included in Table 1. In Table 2, the connected stations of each section, running time, carrying capacity, and maximum and minimum number of marshalling cars in each section are given. With regard to the running time in each section, the operation time in the connected stations is taken into account and that of the two directions is consistent. As for the carrying capacity of each section, it just means the allowed capacity taken up by express freight trains. Supposing the time of original departure and final arrival is 4 hours, the transit time of each demand can be obtained by summing up the running time on all the sections in the candidate path, the original departure, and the final arrival time.

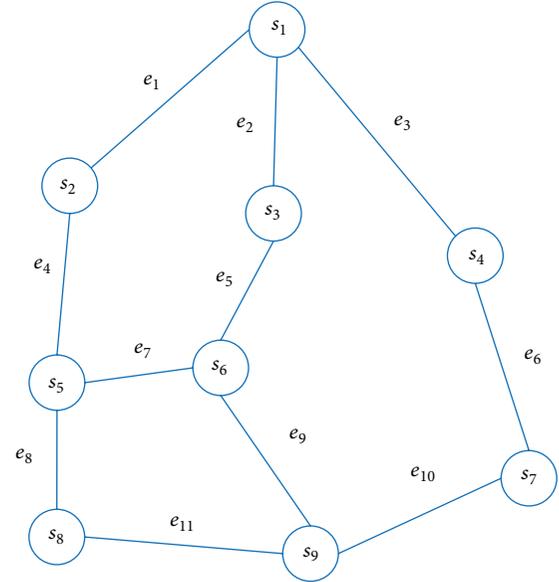


FIGURE 6: Structure of railway physical network.

TABLE 1: Freight demands and their candidate paths.

Demand $OD$	Demand volume	Desired service frequency	Desired transit time	Candidate paths
$s_1-s_5$	25	1	15	$s_1-s_2-s_5$ $s_1-s_3-s_6-s_5$
$s_1-s_6$	60	1	13	$s_1-s_3-s_6$ $s_1-s_2-s_5-s_6$ $s_1-s_3-s_6-s_9$
$s_1-s_9$	78	2	20	$s_1-s_2-s_5-s_8-s_9$ $s_1-s_4-s_7-s_9$
$s_2-s_8$	50	2	18	$s_2-s_5-s_8$
$s_3-s_5$	24	1	12	$s_3-s_6-s_5$ $s_3-s_1-s_2-s_5$
$s_4-s_9$	26	1	15	$s_4-s_7-s_9$ $s_4-s_1-s_3-s_6-s_9$
$s_5-s_9$	61	2	14	$s_5-s_8-s_9$ $s_5-s_6-s_9$
$s_6-s_7$	46	1	16	$s_6-s_9-s_7$ $s_6-s_3-s_1-s_4-s_7$

Furthermore, the maximum departure and arrival capacity of all the stations are 5 trains.

The proposed method is applied to solve the numerical example. The size of the population is 30 and the number of total iteration is 200. The crossover probability is 0.95 and the mutation probability is 0.15. After iterating 116 generations, the satisfactory solution was obtained. The objective function value is 363, which means that 363 cars of freight are transported and the satisfied degree of shippers' transport volume demand is up to 99.73 percent. For clarity, Table 3 is adopted to denote the operation plan of express freight trains, including origin and destination stations, running

TABLE 2: Running time, carrying capacity, and requirement number of train marshalling cars of each section.

Section	Connected stations	Running time	Carrying capacity	Maximum and minimum number of marshalling cars
$e_1$	$s_1, s_2$	6	5	35/25
$e_2$	$s_1, s_3$	4	5	35/25
$e_3$	$s_1, s_4$	7	5	35/25
$e_4$	$s_2, s_5$	4	5	35/25
$e_5$	$s_3, s_6$	3	5	30/20
$e_6$	$s_4, s_7$	4	5	30/20
$e_7$	$s_5, s_6$	2	5	30/20
$e_8$	$s_5, s_8$	3	5	30/20
$e_9$	$s_6, s_9$	4	5	30/20
$e_{10}$	$s_7, s_9$	5	5	30/20
$e_{11}$	$s_8, s_9$	4	5	30/20

TABLE 3: Operation plan of express freight trains.

Origin and destination stations of trains	Running path	Number of marshalling cars	Service frequency
$s_1-s_5$	$s_1-s_2-s_5$	25	1
$s_1-s_6$	$s_1-s_3-s_6$	30	2
$s_1-s_9$	$s_1-s_3-s_6-s_9$	26	3
$s_2-s_8$	$s_2-s_5-s_8$	25	2
$s_3-s_5$	$s_3-s_6-s_5$	24	1
$s_4-s_9$	$s_4-s_7-s_9$	26	1
$s_5-s_9$	$s_5-s_8-s_9$	30	2
$s_6-s_7$	$s_6-s_9-s_7$	23	2

path, number of marshalling cars, and the service frequency of each train.

The test result shows that there are 8 origin and destination pairs to run express trains, the service frequency of which are from 1 to 3, and fourteen express freight trains are operated to transport the demands. The number of marshalling cars for these trains is between 23 and 30. Although some car flows do not choose the shortest path according to the running time, shipper demands on service frequency and transit time are ensured; meanwhile, all capacity constraints are also met.

### 5. Conclusion

Organizing railway express freight transportation according to shipper demands has great significance for improving the market competence of railway freight transportation, as well as promoting the rational division of transportation modes. This paper proposed an optimization method for the operation plan of express freight trains based on shipper demands. An integer programming model considering shipper demands on transport volume, service frequency, and transit time was established under flexible marshalling

condition. A heuristic genetic algorithm with bi-layer coding method that can respond to the characteristics of the model and suit the problem of large scale network was designed. Finally, a numerical example was applied to verify the effectiveness of the model and the algorithm. The result showed that the proposed method can solve the operation plan of single group express freight trains, while the solving time can satisfy the requirement of railway operation. As a matter of fact, multigroup express freight trains are also in operation; thus, the operation problems of these trains will also be studied. Moreover, operation optimization of express freight trains for different kinds of goods and several grades of trains will also be considered to expand the scope of application.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This work was supported by National Natural Science Foundation Project of China under Grant nos. 61263027 and 61104175 and Young Scholars Science Foundation of Lanzhou Jiaotong University under Grant no. 2012031.

### References

- [1] E. Savesberg, *Innovation in European Freight Transportation: Basics, Methodology and Case Studies for European Markets*, Springer, Berlin, Germany, 2008.
- [2] J. A. Pazour, R. D. Meller, and L. M. Pohl, "A model to design a national high-speed rail network for freight distribution," *Transportation Research A*, vol. 44, no. 3, pp. 119–135, 2010.
- [3] B. L. Lin, S. N. Zhu, and S. W. He, "The optimal model of the direct train formation plan for loading area," *China Railway Science*, vol. 16, no. 2, pp. 117–120, 1995.
- [4] T. Crainic, J. Ferland, and J. Rousseau, "A tactical planning model for rail freight transportation," *Transportation Science*, vol. 18, no. 2, pp. 165–184, 1984.
- [5] A. Marín and J. Salmerón, "Tactical design of rail freight networks, part I: exact and heuristic methods," *European Journal of Operational Research*, vol. 90, no. 1, pp. 26–44, 1996.
- [6] M. Yaghini, M. Seyedabadi, and M. M. Khoshraftar, "A population-based algorithm for the railroad blocking problem," *Journal of Industrial Engineering International*, vol. 8, pp. 1–11, 2012.
- [7] D. R. Martinelli and H. Teng, "Optimization of railway operations using neural networks," *Transportation Research C*, vol. 4, no. 1, pp. 33–49, 1996.
- [8] T. G. Crainic, "Service network design in freight transportation," *European Journal of Operational Research*, vol. 122, no. 2, pp. 272–288, 2000.
- [9] C. Lin and S. Chen, "The hierarchical network design problem for time-definite express common carriers," *Transportation Research B*, vol. 38, no. 3, pp. 271–283, 2004.
- [10] A. Ceselli, M. Gatto, M. E. Lübbecke, M. Nunkesser, and H. Schilling, "Optimizing the Cargo Express service of Swiss Federal Railways," *Transportation Science*, vol. 42, no. 4, pp. 450–465, 2008.

- [11] A. Ben-Tal, A. Nemirovski, and M. F. Gorman, "Santa Fe railway uses an operating-plan model to improve its service design," *Interfaces*, vol. 28, no. 4, pp. 1–12, 1998.
- [12] Y. Wang and J. Liu, "Research on the optimization method of railway parcel express trains plan," *Journal of Transportation System Engineering Information Technology*, vol. 7, no. 3, pp. 125–129, 2007.
- [13] C. A. Yano and A. M. Newman, "Scheduling trains and containers with due dates and dynamic arrivals," *Transportation Science*, vol. 35, no. 2, pp. 181–191, 2001.
- [14] B. Crevier, J. Cordeau, and G. Savard, "Integrated operations planning and revenue management for rail freight transportation," *Transportation Research B*, vol. 46, no. 1, pp. 100–119, 2012.
- [15] Y. Guo, S. He, and B. Wang, "Research on optimization of scheduled transit train operating plan," *Journal of the China Railway Society*, vol. 33, no. 5, pp. 8–13, 2011.
- [16] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Optimization*, John Wiley & Sons, New York, NY, USA, 2000.
- [17] H. M. Niu and X. S. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transportation Research C*, vol. 36, pp. 212–230, 2013.
- [18] H. M. Niu and M. H. Zhang, "An optimization to schedule train operations with phase-regular framework for intercity rail lines," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 549374, 13 pages, 2012.

## Research Article

# Development of a Safety Evaluation Model for Provincial Highway

Li Yuan,<sup>1</sup> He-wei Yuan,<sup>2</sup> Yong-feng Ma,<sup>3</sup> and Ying-wei Ren<sup>4</sup>

<sup>1</sup> College of Civil and Transportation Engineering, Hohai University, Nanjing 210098, China

<sup>2</sup> Department of Vehicle, Henan Communication Vocational Technology College, Zhengzhou 450002, China

<sup>3</sup> School of Transportation, Southeast University, Nanjing 210096, China

<sup>4</sup> College of Civil Engineering and Architecture, Shan Dong University of Science and Technology, Qingdao 266510, China

Correspondence should be addressed to Li Yuan; yuanlibox@sina.com

Received 10 November 2013; Revised 22 December 2013; Accepted 24 December 2013; Published 8 January 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Li Yuan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Provincial highway safety is one of the most important issues in transportation. To evaluate or assess the safety performance of provincial highway, traffic crash analysis or traffic conflict analysis has been used for a long history. However, it is very difficult and time-consuming to obtain historical crash data or traffic conflict data. This study analyzes the provincial highways' accident data during 2006–2010, and the characteristics of provincial highway have been investigated; in addition the influencing elements are identified. A comprehensive approach is introduced to evaluate provincial highway safety performance and corresponding models are developed considering the accidents, geometrics, facilities, and traffic environment. The approach will also result in a safety index to indicate the safety performance level of the provincial highway. In this paper, the approach (called safety evaluation approach) is practically applied to evaluate the safety performance of some provincial highways in Hebei Province. Results from the real application indicate that the approach has good applicability and can be used by field safety engineers.

## 1. Introduction

Road safety has become a priority field worldwide and one of the major factors describing the state of the transport system with its positive and negative changes [1], especially provincial highway safety, which is one of the most important issues to be resolved by traffic engineers. According to past statistics, in China, about 32% of traffic crashes and 34% of fatal traffic crashes happen on the provincial highway [2]. These statistical data indicate that provincial highway is the place of significant safety concerns. There is a need to have a practically feasible way to evaluate provincial highway safety performance specifically.

In order to improve highway traffic safety performance, many research studies have been done. The most popular methods to analyze traffic safety performance are based on traffic crash analysis methods, such as crash frequency, crash rate, crash severity, and crash statistic models [3, 4]. However, all these crash-based analyses are based on past historical

crash data. Accumulating crash data could be a problem to many researchers, mainly due to the difficulty to obtain these data and the reliability of past crash data [5]. For example, to determine the black spots with high traffic crash rates in China, it usually needs to take about 3 to 10 years to accumulate traffic crash data. In fact, within the time period, land use situation could change, resulting in biased safety improvement decisions.

With such considerations, noncrash-based analysis methods have been used by many researchers [6]. One of the methods is traffic conflict analysis (TCA) [7]. Based on the definition from Perkins, a traffic conflict is defined as the interactions between vehicles and such interactions would result in some actions, such as braking and movement direction change [8]. TCA methods generally cover conflicting points, number of conflicts, conflict rate, conflict distribution, and conflict forecasting models [9, 10]. This method has been used by many researchers to analyze traffic safety

performance. For example, in USA, TCA has been used to evaluate roadway intersection traffic safety. In Britain, TCA has been used to evaluate the impacts of safety improvement countermeasures. In Sweden, TCA is used to evaluate safety performance of high-speed highways and corresponding safety countermeasures. Besides, TCA has been widely used in Japan and Austria. In China, Lu et al. [11] have taken the weighted sum of the crossing point numbers of ideal movement trajectories as the basic conflict model for the safety level evaluation of highway intersections. In this model, the physical conditions are used to assess safety levels without the need for crash data. However, it is still hindered by some limitations for signalized intersection. Zhou et al. [12] established a conflict hazardous assessment model (CHAM) for the evaluation of urban intersection safety incorporating factors such as conflict types, conflict angles, velocity, and weight; this model is mainly used for urban intersections. In recent years, traffic conflict simulation models have been attempted to analyze traffic safety performance at roadway intersections [13, 14]. Traffic conflict simulation model development needs long-term data accumulation and simulation model parameters could change as environmental conditions change. In addition, the judgment and determination of traffic conflicts are more subjective nature. Different observers may give different traffic conflict judgments. This limits its applicability in real situation [15].

As discussed above, both crash-based and conflict-based safety analysis methods have some limitations in real applications. Highway safety evaluation is not implemented on a systematic and integrated appraisal method but simply selects a spot or section chosen by accident or conflict. The method could miss the possible accident causation by typical road condition and various traffic characteristics. Even though the traffic accident did not occur on a certain road during past few years, still it is possible to occur on that road in the near future.

To maximize traffic safety improvement project, in the research, another approach called safety evaluation model is attempted. This approach is a comprehensive evaluation method based on accidents, geometrics, facilities, and traffic environment. With such characteristics, this approach could give a real evaluation to provincial highway safety performance. Particularly, the type of method has advantages, such as low cost, high efficiency, and less time-consuming. Furthermore, this approach can be better used by field safety engineers to find potential safety problems and corresponding countermeasures can be implemented so that possible traffic safety problems can be prevented. To develop the model, the related data of 40 provincial highway (total 4,100 Km as of years of 2010–2012) and accident data (years of 2006–2010: 3 years) in China are collected.

## 2. Approach

The first step was getting a group of the accident features of the region through the analysis on general status and accidents. Elements for evaluating hazardous highways were collected through investigating documents from in and out

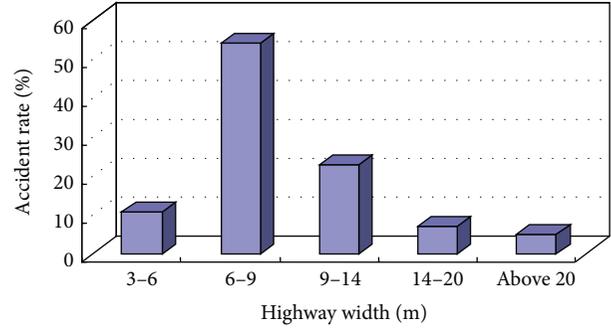


FIGURE 1: Ratio of accidents according to highway width.

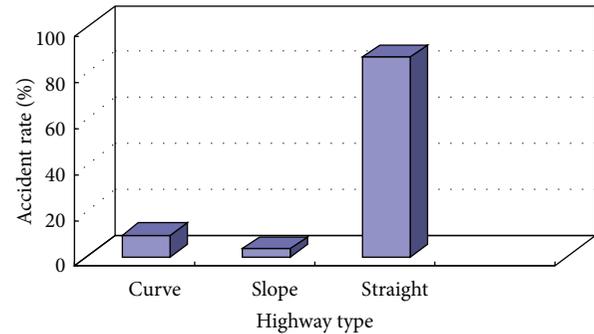


FIGURE 2: Ratio of accidents according to highway type.

of the country. Then through AHP (analysis of hierarchy process) method, relative weight of collected elements was calculated. Based on the evaluation model through this process, the hazardous highway selection criteria and scoring method were prepared.

*2.1. Analysis on Accidents in Provincial Highways.* Among the accident data provided by the Police Agency, only the data happening on the provincial highways (years of 2006–2010) were collected to analyze the features of accidents. Analyzed accident features were used as variables for provincial highway evaluation index along with collected geometrical features and environmental factors through the literature study.

*2.1.1. Accident Feature according to Geometrical Structures of a Highway.* As a result of comparing accident rates in provincial highway with width of the highway, it was shown that the accidents occur most frequently in highways having 6–9 m width. It shows that the alignment and safety facilities are relatively inappropriate when the width of highway is narrow, as shown in Figure 1.

In terms of the accident rate according to highway type, highway had high accident rate in a straight section with long expansion length (shown in Figure 2). However, the fatal rate compared to accident case was relatively higher in the curve and slope sections, as shown in Figure 3, proving that curve and slope sections have higher risk than straight section.

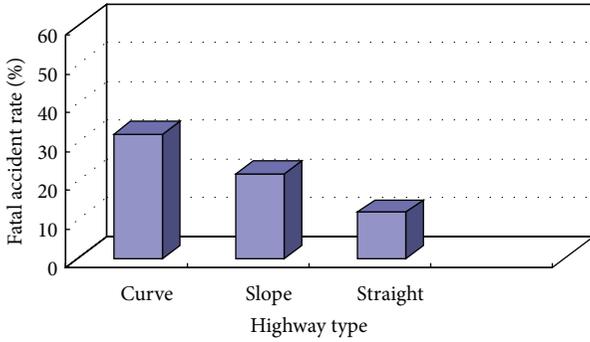


FIGURE 3: Fatal rate according to highway type.

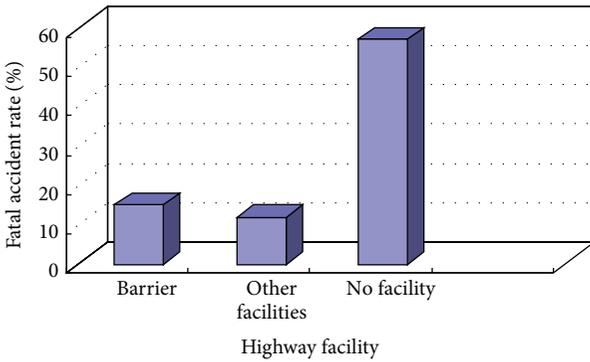


FIGURE 4: Fatal rate according to highway traffic facility.

2.1.2. *Accident Feature according to Traffic Facilities of a Highway.* In terms of the fatal accident rate according to the highway traffic facilities, these sections with no traffic facilities had higher risk, as shown in Figure 4, proving that no-facilities section is in serious condition.

2.1.3. *Accident Feature according to Accident Types.* In terms of the fatal accident rate of vehicle to pedestrian, vehicle/vehicle and single vehicle accidents, as shown in Figure 5, the order of accident types is vehicle, to pedestrian, vehicle/vehicle, and single vehicle accident, proving how the walking environment in provincial highways is in serious condition.

2.1.4. *Results of Analysis.* Accidents and highway features of provincial highways through the analysis are as follows: (1) the ratio of accident occurring in the width of the highway ranks highest in a width of 6–9 m; (2) provincial highways have higher fatal rate in curve and slope sections compared to straight section; (3) the order of fatal rate in provincial highway was proven to be vehicle to pedestrian, vehicle/vehicle, and single vehicle accident.

2.2. *Contributing Factors to Provincial Highway Safety.* There are many factors that could affect provincial highway safety performance, including subjective factors and objective factors. Subjective factors mainly refer to human-related factors such as driving behavior. Objective factors are mainly related

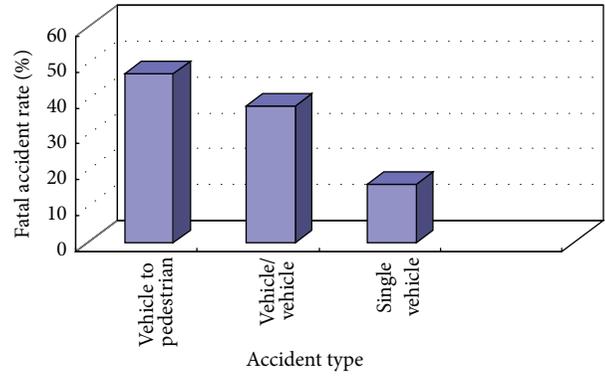


FIGURE 5: Fatal rate according to highway accident type.

to the physical characteristics of highway, such as geometrics, traffic signs and markers, and traffic characteristics. In real applications, it is hard to measure subjective factors. Thus, only objective factors are considered in developing models for safety evaluation.

According to the past research, including analysis accident feature of provincial highway, conference survey, focus group discussion, field observations, and letter survey, the major contributing factors include accident severity, geometric feature (horizontal curve, sight distance, longitudinal slope, width of the highway, and shoulder), traffic facilities (traffic signs and marking, guiding facilities, and safety facilities), traffic environment (traffic volume and configuration, crosswalk).

### 3. Methodology

As stated previously, provincial highway safety evaluation should be not only based on accident severity but also based on geometric feature, traffic facility, and traffic environment. Before the final model specifications were decided, statistical analysis (such as principle component analysis) was performed to identify significant variables for the models. All the variables included in the models (shown in Figure 6) were significant. Some variables such as lighting condition and parking, were not included in the models due to the insignificance.

### 4. Evaluation Modeling Procedure

As discussed previously, to evaluate provincial highway safety evaluation, quantified indices are needed. As seen in Figure 6, the main indices for geometric feature ( $F_2$ ), traffic facility ( $F_3$ ), and traffic environment ( $F_4$ ) are horizontal curve ( $F_{21}$ ), sight distance ( $F_{22}$ ), longitudinal slope ( $F_{23}$ ), lane width ( $F_{24}$ ), shoulder width ( $F_{25}$ ), traffic sign ( $F_{31}$ ), traffic marking ( $F_{32}$ ), guiding facility ( $F_{33}$ ), safety facility ( $F_{34}$ ), crosswalk ( $F_{41}$ ), traffic volume ( $F_{42}$ ), and heavy vehicle ratio ( $F_{43}$ ), respectively. Conceptually, each area (geometric, traffic facility, and traffic environment) should have a summarized index to quantify the safety of provincial highway in the corresponding area. This summarized index could be a linearly weighted

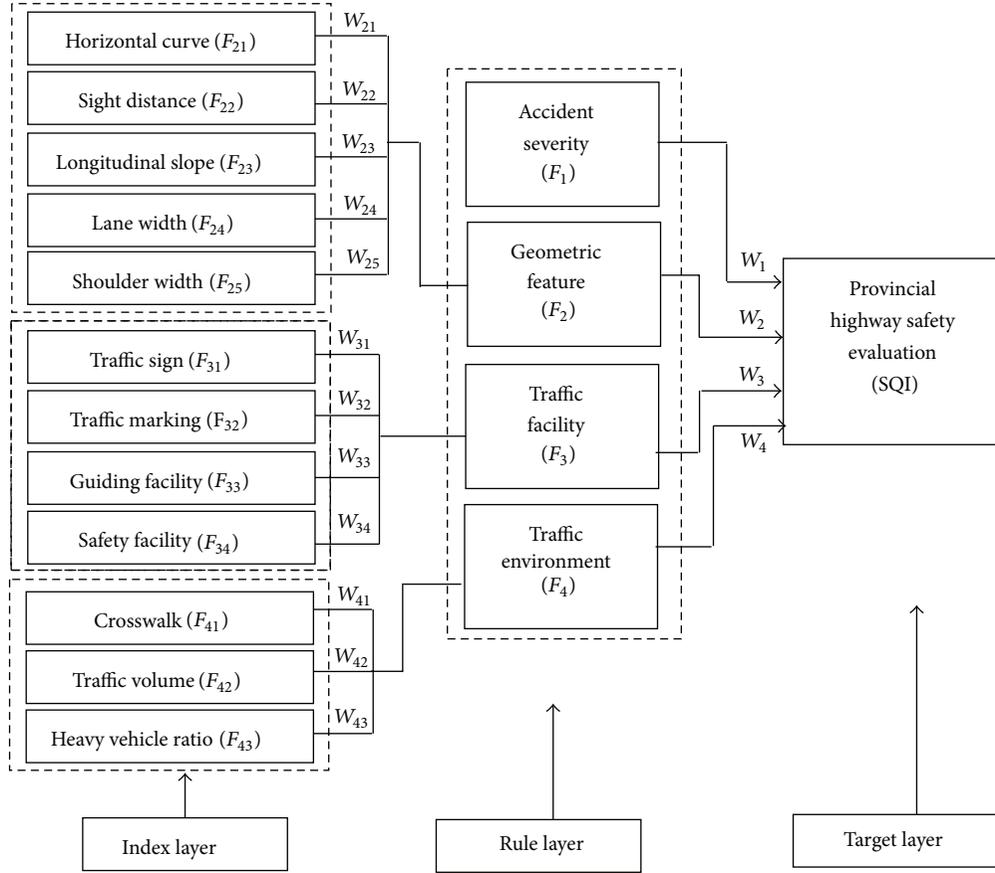


FIGURE 6: Logic structure for provincial highway safety evaluation.

combination of safety evaluation. According to Figure 6, the following equations could be used:

$$\begin{aligned}
 F_2 &= W_{21}F_{21} + W_{22}F_{22} + W_{23}F_{23} + W_{24}F_{24} + W_{25}F_{25}, \\
 F_3 &= W_{31}F_{31} + W_{32}F_{32} + W_{33}F_{33} + W_{34}F_{34}, \\
 F_4 &= W_{41}F_{41} + W_{42}F_{42} + W_{43}F_{43},
 \end{aligned} \quad (1)$$

where  $F_2$ ,  $F_3$ , and  $F_4$  are summary indices for geometric feature, traffic facility, traffic environment, respectively, and  $W_{ij}$  ( $i = 2, 3, 4$ , and  $j = 1, 2, 3, 4, 5$ ) are corresponding weights to be estimated through the modeling procedure.

To evaluate the overall safety performance ( $F$ ) of the provincial highway,  $F_1$ ,  $F_2$ ,  $F_3$ , and  $F_4$  should also be linearly combined with weights  $W_1$ ,  $W_2$ ,  $W_3$ , and  $W_4$ .

Consider

$$F = W_1F_1 + W_2F_2 + W_3F_3 + W_4F_4, \quad (2)$$

where  $F$  is the index which quantifies the overall safety performance (quality) of the provincial highway.

**4.1. Determination of Index Weights.** To calibrate the models shown in (1) to (2), AHP method was used to determine weights. It is not only to be applied but also to be widely used in decision-making process of public and official areas

because of its accurate theoretical evidence. Prior to performing AHP method, a survey was conducted on 32 experts, including researchers who have experts in the relevant areas, transportation engineers working in the similar areas, traffic police officers, and roadway motorized vehicle users. Many of the interviewees had experienced traffic conflicts either as a driver or as a vulnerable user. People involved in the survey were required to fill a survey form to give a rating of the importance of each variable as relative to other variables, independently. Detailed explanations were presented before the forms were completed. After all the survey forms were completed, the AHP method was used to obtain the weights included in (1) to (2).

**4.1.1. Weights of Rule Layer.** Using AHP method, the weight of accident severity, geometric features, traffic facility, and traffic environment had been calculated. Table 1 is the result of calculating the weight for evaluators.

Geometric mean and weight about accident severity can be computed as follows.

$$(i) \text{ Geometric Mean} = (1 \times 2 \times 3 \times 5)^{(1/4)} = 2.34.$$

$$(ii) \text{ Weight} = 2.34/4.91 = 0.47.$$

Similar to analysis mentioned above, values of other geometric means and weights are shown in Table 1.

TABLE 1: Result of calculating the weight for evaluators.

	Accident severity	Geometric feature	Traffic facility	Traffic environment	Geometric mean	Weight
Accident severity	1	2	3	5	2.34	0.47
Geometric feature	0.5	1	2	4	1.41	0.28
Traffic facility	0.3333	0.5	1	2	0.76	0.15
Traffic environment	0.2	0.25	0.5	1	0.4	0.10
Sums	—	—	—	—	4.91	1

TABLE 2: Result of calculating constancy index of evaluators.

	Accident severity	Geometric feature	Traffic facility	Traffic environment	Row sum	Row sum/weight
Accident severity	0.47	0.56	0.45	0.5	1.98	4.16
Geometric feature	0.28	0.28	0.3	0.4	1.26	4.38
Traffic facility	0.15	0.14	0.15	0.2	0.64	4.14
Traffic environment	0.09	0.08	0.07	0.1	0.34	4.2
Sums	—	—	—	—	—	16.9

4.1.2. *Test AHP Result.* To test AHP result, weights were multiplied by each column of judgment standard in the comparison table. The driven figures are summarized in Table 2. Then the sum of each column was divided into the weights of each rule layers.

After computing the means of values obtained from Table 2, consistency index had been calculated as follows.

- (i) Means =  $16.9/4 = 4.22$ .
- (ii) CI (Conformity) =  $(\text{Means} - n)/(n - 1) = (4.22 - 4)/(4 - 1) = 0.07$ .
- (iii) CR (Consistency Index) =  $CI/RI = 0.07/0.9 = 0.077$ ,

where: RI is Random Index that can be obtained from table according to the number of factors ( $n$ ), and its value is 0.9.

As a result of calculating consistency index (CR), its value is 0.077; when it is less than 0.1, it is judged that the process of decision making is consistent.

Similar to analysis mentioned above, weights included in (1) to (2) are determined and summarized in Table 3. The following equation was obtained based on the AHP method:

$$\begin{aligned}
 F_2 &= 0.23F_{21} + 0.21F_{22} + 0.31F_{23} + 0.13F_{24} + 0.12F_{25}, \\
 F_3 &= 0.3F_{31} + 0.13F_{32} + 0.2F_{33} + 0.37F_{34}, \\
 F_4 &= 0.32F_{41} + 0.45F_{42} + 0.23F_{43}, \\
 F &= 0.47F_1 + 0.28F_2 + 0.15F_3 + 0.1F_4.
 \end{aligned}
 \tag{3}$$

## 5. Setting of Evaluation Criteria for Evaluation Index

5.1. *Evaluation Criteria of Accident Severity.* Method considering the accident severity is applied by calculating the index

“traffic accident.” Traffic accident index is calculated using the following equation:

$$AS_i = \left( \frac{A_i}{\sum_{i=1}^n A_i} + \frac{PS_i}{\sum_{p=1}^n PS_i} \right) \times 100\%, \tag{4}$$

$$PS_i = \sum P_j S_j, \tag{5}$$

where  $AS_i$  is the accident severity index,  $A_i$  is the number of accident in  $i$  section,  $PS_i$  is the accident severity in  $i$  section,  $P_j$  is the weighting factor of person’s accident, and  $S_j$  is the number of relating person’s accidents.

In case of the weighting factor, its value is added for each type of accidents by categorizing accidents with human damage into death, serious, and slight injuries. Values for adding weight were death, serious injury, and slight injury = 28 : 3 : 1 [16].

Accident severity index ( $AS_i$ ) calculated using (4) and (5) was summarized in an order to be scored according to the comparative evaluation for subjected sections (shown in Table 4).

### 5.2. Evaluation Criteria according to Geometric Features

5.2.1. *Horizontal Curve.* The risk of accident in horizontal curve has been proven through the analysis on the features of accidents on provincial highway [17]. The radius of horizontal curve in provincial highway can be installed up to 115–360 m according to the design speed and is scored to the relative evaluation for each subjected section. Table 5 is the radius standard according to the design speed.

Horizontal curve standard for provincial highway is general radius for ground level and min. radius for mountain level;

- (i) calculation method: number of understandard radii/section lengths (km)
- (ii) scoring method: the calculated value is summarized in the order to be scored by relative evaluation for each subjective section (shown in Table 6).

TABLE 3: Weights of comprehensive evaluation on provincial highway.

Index	Rule layer	Weight	Subindex	Index layer	Combined weight
Accident severity	$W_1$	0.47		—	
Geometric feature	$W_2$	0.28	Horizontal curve	$w_{12}$	0.23
			Sight distance	$w_{22}$	0.21
			Longitudinal slope	$w_{23}$	0.31
			Lane width	$w_{24}$	0.13
			Shoulder width	$w_{25}$	0.12
Traffic facility	$W_3$	0.15	Traffic sign	$w_{31}$	0.3
			Traffic marking	$w_{32}$	0.13
			Guiding facility	$w_{33}$	0.2
			Safety facility	$w_{34}$	0.37
Traffic environment	$W_4$	0.10	Crosswalk	$w_{41}$	0.32
			Traffic volume	$w_{42}$	0.45
			Heavy vehicle ratio	$w_{43}$	0.23

TABLE 4: Scoring method according to accident severity index ( $AS_i$ ).

Relative ratio ( $AS_i$ )	Under 20%	20%–40%	40%–60%	60%–100%
Score	$\leq 40$	$\leq 60$	$\leq 80$	$\leq 100$

TABLE 5: Radius standard for plane curve for design speed.

Classification	Unit	Design speed (km/h)		
		100	80	60
Radius (general)	m	700	400	200
Min. radius		360	220	115

TABLE 6: Scoring method according to number of curve sections.

Relative ratio	Under 10%	10%–20%	20%–30%	30%–50%
Score	$\leq 40$	$\leq 60$	$\leq 80$	$\leq 100$

TABLE 7: Minimum sight distance standard according to design speed.

Design speed (km/h)	100	80	60
Sight distance (m)	160	110	75

TABLE 8: Scoring method according to number of sight distance defect sections.

Relative ratio	Under 10%	10%–20%	20%–30%	30%–50%
Score	$\leq 40$	$\leq 60$	$\leq 80$	$\leq 100$

TABLE 9: Longitudinal slope standard for design speed.

Design speed (km/h)	100	80	60
Maximum longitudinal slope	4%	5%	6%

5.2.2. *Sight Distance.* Sight distance defect is the major cause of accidents such as previous collision and others frequently occurring in provincial highway [18, 19] the score should be

TABLE 10: Scoring method according to number of longitudinal slope sections.

Relative ratio	Under 10%	10%–20%	20%–30%	30%–50%
Score	$\leq 40$	$\leq 60$	$\leq 80$	$\leq 100$

TABLE 11: Scoring method according to mean lane width.

Lane width (m)	3.5–3.75	3.25–3.5	3–3.25	Under 3
Score	$\leq 40$	$\leq 60$	$\leq 80$	$\leq 100$

high if the ratio of sight distance defect section is high. Sight distance defect section refers to the standard less than the minimum sight distance standard according to the design speed (shown in Table 7);

- (i) calculation method: number of sight distance defect sections/section lengths;
- (ii) scoring method: the calculated value is summarized in the order to be scored by relative evaluation for each subjective section (shown in Table 8).

5.2.3. *Longitudinal Slope.* The risk of accident in longitudinal slope has been proven through the preceding researches. The longitudinal slope in provincial highway can be installed up to 4%–6% according to the design speed and is scored to the relative assessment for each subjected section. Table 9 is the maximum longitudinal slope standard according to the design speed;

- (i) calculation method: number of understandard longitudinal slopes/section lengths (km);
- (ii) scoring method: the calculated value is summarized in the order to be scored by relative evaluation for each subjective section (shown in Table 10).

5.2.4. *Lane Width.* It has been proven that the risk of accidents is higher with narrow lane width according to the

TABLE 12: Calculation and scoring method according to assessment items.

Index	Subindex	Assessment element	Standard	Score
Geometric feature ( $F_2$ )	Horizontal curve ( $F_{21}$ )	Number of curve sections (%)	Under 10	$\leq 40$
			10–20	$\leq 60$
			20–30	$\leq 80$
			30–50	$\leq 100$
			Under 10	$\leq 40$
	Sight distance ( $F_{22}$ )	Number of sight distance defect sections (%)	10–20	$\leq 60$
			20–30	$\leq 80$
			30–50	$\leq 100$
	Longitudinal slop ( $F_{23}$ )	Number of longitudinal slope sections (%)	Under 10	$\leq 40$
			10–20	$\leq 60$
			20–30	$\leq 80$
			30–50	$\leq 100$
			3.5–3.75	$\leq 40$
	Lane width ( $F_{24}$ )	Mean lane width (m)	3.25–3.5	$\leq 60$
			3–3.25	$\leq 80$
Under 3.25			$\leq 100$	
Shoulder width ( $F_{25}$ )	Mean of shoulder width (m)	Nonpavement	100	
		0–1.5	Interpolation	
		Above 1.5	—	

TABLE 13: Survey form for traffic facility.

Index	Subindex	Items contributing to the rating	Score range
Traffic facility ( $F_3$ )	Traffic sign ( $F_{31}$ )	The installation meets the standard, good visibility, and reasonable message	$\leq 40$
		Poor visibility and reasonable message and installation	$\leq 60$
		Worse visibility and some unseasonable message and installation	$\leq 80$
		Worst visibility and most unreasonable message and installation	$\leq 100$
		The installation meets the standard and good visibility	$\leq 40$
	Traffic marking ( $F_{32}$ )	Poor visibility and reasonable installation	$\leq 60$
		Worse visibility and some unseasonable installation	$\leq 80$
		Worst visibility and most unreasonable installation	$\leq 100$
	Guiding facility ( $F_{33}$ )	Spacing setting, number, and layout size satisfy standard	$\leq 40$
		Spacing setting, number, and layout size basically satisfy standard	$\leq 60$
		Most spacing setting and number don not satisfy the standard. Layout size is slightly smaller than the normative value.	$\leq 80$
		Spacing setting and number don not satisfy the standard. Layout size is less than the minimum.	$\leq 100$
	Safety facility ( $F_{34}$ )	Continuous safety barriers or 10 m safety zone	$\leq 40$
		Safety barriers for hard obstacles or steep slopes within 6 m	$\leq 60$
		Safety barriers for more than half of hard obstacles or steep slopes within 3 m	$\leq 80$
		Barriers not generally used	$\leq 100$

preceding researches. Therefore, more score is given to lane width narrower lane width. The lane width standard is 3.0 m–3.5 m according to the design speed. Calculation method according to the lane width is mean lane width of subjected sections, and Table 11 is the scoring method according to mean lane width.

5.2.5. *Shoulder Width.* Shoulder can be used to install safety facilities or as pedestrian area, and correlation with accident had been proven by preceding researches [20]. Therefore, score is granted more at narrower shoulder width. Shoulder width under ideal of provincial highway is 1.5 m or more. In case of unpaved shoulders, pavement is serious matter.

TABLE 14: Calculation and scoring method according to assessment items.

Index	Subindex	Assessment element	Standard	Score
Traffic environment ( $F_4$ )	Crosswalks ( $F_{41}$ )	Number of crosswalk in section/section length	Under 10%	$\leq 40$
			10%–20%	$\leq 60$
			20%–30%	$\leq 80$
			Above 30%	$\leq 100$
	Traffic volume ( $F_{42}$ )	ADT per lane	Lower traffic	$\leq 40$
			Midtraffic	$\leq 60$
Heavy vehicle ratio ( $F_{43}$ )	Heavy vehicle/traffic volume	Heavier traffic	$\leq 100$	
		Below 36%	Interpolation	
			Above 36%	100

TABLE 15: Definition of level of provincial highway safety.

Level of safety	Rate	Color	Condition
A	$\leq 40$	Green	Very safe
B	$\leq 60$	Blue	Safe
C	$\leq 80$	Orange	Unsafe
D	$\leq 100$	Red	Dangerous

Table 12 is the summary of calculation and scoring methods of each geometric feature presented above.

5.3. *Evaluation Criterion of Traffic Facility.* Similar to the analysis of method presented above, Table 13 is the survey form designed and revised in the research. This form provides the rating score range for each evaluation index and corresponding of traffic facility. Field surveyors need to survey the traffic facilities and fill the corresponding parts in the form. It is easy to use the form although some subjective effects could affect rating results.

5.4. *Evaluation Criterion of Traffic Environment*

5.4.1. *Crosswalk.* Provincial highway has high accident rate during pedestrian’s crossing of the highway [21]. Therefore, considering the ratio of the crosswalk, higher score is given to higher number of crosswalks;

- (i) calculation method: number of crosswalks in section/section lengths;
- (ii) scoring method: Table 14 is the scoring method according to the number of crosswalks in a section.

5.4.2. *Traffic Volume.* Higher traffic volume refers to higher risk of accidents [22]. Therefore, highway with higher traffic volume are called hazardous highway;

- (i) calculation method: using the ADT (vehicles/day) of each lane of subjected section;
- (ii) scoring method: Table 14 showed the scoring method according to the volume. ADT had been calculated as shown below;
- (iii) low traffic:  $ADT < \mu - \sigma$  of subjected section;

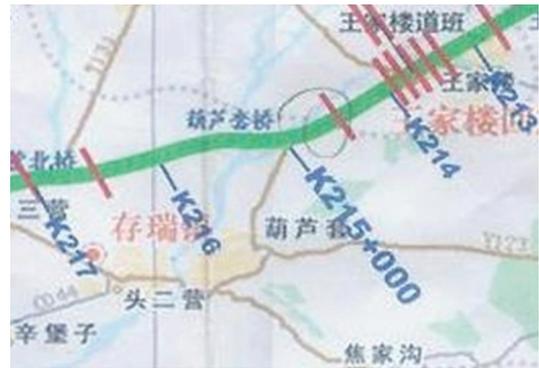


FIGURE 7: Hebei provincial highway S241 sections.

- (iv) mid traffic:  $\mu - \sigma < ADT < \mu + \sigma$  of subjected section;
- (v) heavy traffic:  $ADT > \mu + \sigma$  of subjected section;

where  $\mu$  is the mean ADT for number of lanes and  $\sigma$  is the distribution.

5.4.3. *Heavy Vehicle Ratio.* Higher rate of heavy vehicle is known to increase the number of accidents. Therefore, higher score is given to the section with higher ratio of heavy vehicles. According to the traffic volume statistical report, the mean of daily heavy vehicle in provincial highway is 36% (shown in Table 14).

**6. Usage of Overall Safety Rating of Provincial Highway**

Safety rating in use either predicts safety outcomes for given designs or provides a retrospective assessment. For a given provincial highway, SQI value (overall rating) calculated from previous equations is used as the base for level of safety. SQI value can be classified into 4 levels from A to D with A indicating the best safety service and D the worst safety service. In order to define level of safety service, field survey activities to many provincial highways were conducted in Hebei Province, Jiangsu Province, Shandong Province, and other areas in China. Based on field survey results, the level of provincial highway safety is shown in Table 15.

TABLE 16: Survey results of Hebei provincial highway S241 sections.

Test section	Evaluation index	K213-k214	K215-K216
Geometric feature	Accident severity	70	5
	Horizontal curve	65	70
	Sight distance	70	80
	Longitudinal slope	45	50
	Lane width	20	20
	Shoulder width	45	45
Traffic facility	Traffic sign	50	60
	Traffic marking	30	30
	Guiding facility	65	75
	Safety facility	90	90
Traffic environment	Crosswalk	45	45
	Traffic volume	30	30
	Heavy vehicle ratio	80	80

TABLE 17: Evaluation results of Hebei provincial highway S241 sections.

Test section	Evaluation rating				Overall rating (SQI)	Level
	Accident severity	Geometric feature	Traffic facility	Traffic environment		
1 (k213-k214)	70	68	65	45	66	C*
2 (k215-k216)	5	52	65	46	31	A <sup>#</sup>

\*Orange color and <sup>#</sup>blue color.

For the levels of A and B, there is no need to propose any special improvement activities. Only regular or routine safety management will be sufficient. For the level of C, certain adjustments and minor additions in improvement are needed. For the level of D, major adjustments in improvement are certainly needed.

### 7. Case Study

Some test provincial highway sections in Hebei Province were used for the application purpose. 1000-meter-long sections of provincial highway were shown in Figure 7, and survey scores were listed in Table 16. The results were obtained based on the criteria provided in Table 4 and Tables 12–14.

Based on model (3), the evaluation results were listed in Table 17.

It was shown that the safety level of section two is A that means the best safety service, but traffic facility of section was considered fair, indicating that it is a potential risk factor, and those of the geometric and traffic environments were considered safe. Whereby, minor safety maintenance is required. The overall results of section one illustrated that the safety level is C and the geometric and traffic facility safeties were poor. Therefore, section one needed middle-scale improvement.

### 8. Conclusions

This research proposes a standard and safety evaluation measure for provincial highway. To provide realistic standard,

the data of accidents occurring in provincial highways were analyzed to drive their features. Also by researching previous researchers, major elements were selected to evaluate provincial highway, and the weights of these elements had been calculated using AHP method.

It is judged that using the result of this research can maximize the efficacy of highway safety management project since it can select hazardous highway not only based on the number of accidents but also based on geometric feature, traffic facility, and traffic environment. Also, this research is different from the ones in the past in a way that crash-based analysis and traffic-conflict analysis were used which can be difficult to apply in real cases. On the contrary, this research had made it easy to apply site application.

With models developed in the research project, provincial highway engineers are able to better or more objectively evaluate safety performance so that optimized improvement options could be planned. However, in order to reasonably use the evaluation models, the models should be upgraded on the yearly basis with the new field data collected. In fact, the approach described in the paper can be implemented into a computer-based package with friendly software interfaces. Field engineers can use the computer-based package to perform field surveys and enter data to the package. This can reduce significant amount of time to process the survey results and make the survey work much easier.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

This research was supported in part by the National Natural Science Foundation of China with nos. 51308192 and 51208100. The authors would like to show great appreciation for these supports.

## References

- [1] G. Dell'Acqua and M. Busiello, "Safety data analysis to evaluate highway alignment consistency," *Transportation Research Record*, no. 2349, pp. 121–128, 2013.
- [2] F. Q. Pan, L. X. Zhang, and J. Lu, "A method for determining the number of traffic conflict points between vehicles at major-minor highway intersections," *Traffic Injury Prevention*, vol. 14, no. 4, pp. 424–433, 2013.
- [3] S. Danief and R. Edward, "Variability in rural accident reporting," *Journal of the Transportation Research Board*, no. 910, pp. 8–13, 1983.
- [4] M. Wu and C. Z. Wu, "Methods to improve the plain terrain expanded freeway under accident reason analysis," *Journal of Transport Information and Safety*, vol. 29, no. 6, pp. 73–77, 2011.
- [5] A. Joaquin, L. Griselda, and D. O. Juan, "Analysis of traffic accident severity using decision rules via decision trees," *Expert Systems with Application*, vol. 40, no. 15, pp. 6047–6054, 2013.
- [6] B. Son, M. Park, and S. Lee, "A study for hazardous road selection criteria for provincial roads," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 6, pp. 3426–3440, 2005.
- [7] L. Yuan and J. Lu, "Safety evaluation and improvements for highway intersections," *Journal of the Transportation Research Board*, no. 2060, pp. 46–52, 2008.
- [8] S. R. Perkins, "Traffic conflict characteristics—accident potential at intersections," Highway Research Record 225, HRB, National Research Council, Washington, DC, USA, 1968.
- [9] J. J. Lu, Q. J. Xiang, and S. Dissanayake, "The ways to improve highway safety," *Journal of Advances in Transportation Studies*, vol. 2, pp. 3–14, 2004.
- [10] G. Davis, J. Hourdos, and H. Xiong, "Outline of a causal theory of traffic conflicts and collisions," in *Proceedings of the 87th Annual Meeting*, Washington, DC, USA, 2008.
- [11] J. Lu, F. Q. Pan, and Q. J. Xiang, "Level-of-safety service for safety performance evaluation of highway intersections," *Journal of the Transportation Research Board*, no. 2075, pp. 24–33, 2008.
- [12] S. Zhou, J. Sun, X. An, and K. P. Li, "The development of a conflict hazardous assessment model for evaluating urban intersection safety," *Transport*, vol. 26, no. 2, pp. 216–223, 2011.
- [13] H. E. Ross, D. L. Sicking, and R. A. Zimmer, "Recommended procedures for the safety performance evaluation of highway features," NCHRP Report 350, TRB, National Research Council, Washington, DC, USA, 1993.
- [14] F. Huang and P. Liu, "Identifying if VISSIM simulation model and SSAM provide reasonable estimates for field measured traffic conflicts at signalized intersections," *Accident Analysis and Prevention*, vol. 50, pp. 1014–1024, 2012.
- [15] P. Margie, S. Richard, and S. David, "World report on road traffic injury prevention," Report, World Health Organization, Geneva, Switzerland, 2004.
- [16] K. F. Wu and E. T. Donnell, "Exploring the association between traffic safety and geometric design consistency based on vehicle speed metrics," *Journal of Transportation Engineering*, vol. 139, no. 7, pp. 738–748, 2013.
- [17] F. J. Camacho-Torregrosa, A. M. Pérez-Zuriaga, J. M. Campoy-Ungría, and A. García-García, "New geometric design consistency model based on operating speed profiles for road safety evaluation," *Accident Analysis and Prevention*, vol. 61, pp. 33–42, 2013.
- [18] N. Roussiamanis, A. Kaltsounis, and S. Vardaki, "A review of stopping sight distance in road guidelines," *Proceedings of the Institution of Civil Engineers: Transport*, vol. 166, no. 5, pp. 305–312, 2013.
- [19] S. E. Ibrahim, T. Sayed, and K. Ismail, "Methodology for safety optimization of highway cross-sections for horizontal curves with restricted sight distance," *Accident Analysis and Prevention*, vol. 49, pp. 476–485, 2012.
- [20] Z. Z. Li, K. Kepaptsoglou, and Y. Lee, "Safety effects of shoulder paving for rural and urban interstate, multilane, and two-lane highways," *Journal of Transportation Engineering*, vol. 139, no. 10, pp. 1010–1019, 2013.
- [21] C. Havard and A. Willis, "Effects of installing a marked crosswalk on road crossing behaviour and perceptions of the environment," *Transportation Research F*, vol. 15, no. 3, pp. 249–260, 2012.
- [22] D. Stavrinou, J. L. Jones, and A. A. Garner, "Impact of distracted driving on safety and traffic flow," *Accident Analysis and Prevention*, vol. 61, pp. 63–70, 2013.

## Research Article

# Tour Route Multiobjective Optimization Design Based on the Tourist Satisfaction

**Yan Han, Hongzhi Guan, and Jiaying Duan**

*Key laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China*

Correspondence should be addressed to Yan Han; hanyan422@bjut.edu.cn

Received 20 September 2013; Accepted 4 December 2013; Published 8 January 2014

Academic Editor: Huimin Niu

Copyright © 2014 Yan Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The question prompted is how to design the tour route to make the tourists get the maximum satisfactions considering the tourists' demand. The influence factors of the tour route choices of tourists were analyzed and tourists' behavior characteristics and psychological preferences were regarded as the important influence factors based on the tourist behavioral theories. A questionnaire of tourists' tour route information and satisfaction degree was carried out. Some information about the scene spot and tourists demand and tour behaviors characteristic such as visit frequency, number of attractions visited was obtained and analyzed. Based on the convey datum, tour routes multiobjective optimization functions were prompted for the tour route design regarding the maximum satisfaction and the minimum tour distance as the optimal objective. The available routes are listed and categorized. Based on the particle swarm optimization model, the priorities of the tour route are calculated and finally the suggestion depth tour route and quick route tour routes are given considering the different tour demands of tourists. The results can offer constructive suggestions on how to design tour routes on the part of tourism enterprises and how to choose a proper tour route on the part of tourists.

## 1. Introduction

With the development of economics, more and more people have leisure time to travel. During the travel, it is necessary for the tourist to determine the travel destination and traffic mode and visit more scenic spots and historical sites in the limited time budget and get the maximum satisfactions. For some tourists, it will take them some time to determine the tour route before or during the travel [1], especially for those tourists who visit the scenic spot for the first time. So for tourism enterprises the most important thing is to provide reasonable tour route and different scenic spots to attract more tourists and enhance the attractions charm and popularity. More and more tour route products are analyzed and provided so as to meet the different demands of tourists. And the problem prompted for the tourism enterprises is how to design and provide the tour route system to make the tourists get the maximum satisfactions according to the tourists' demand, what the tourists' demand is, and how to obtain the tourists' demand. The tour route system refers to the continuous space chain connected with each landscape feature point of a tourist area with the concepts of time and

space. It has three different levels: one is accessibility tourism routes connected by several tourist center cities; the second is the main tourist routes in which the tourist center of the city is defined as a "home base" and various tourist scenic spots are links to the tourist center of the city; the third is tour route within the scenic which is discussed in this paper [2].

The design of tour route is associated with the development of Chinese tourism academic research and is divided into three categories. One is from space research, mainly focused on the characteristics and evolution of the spatial structure of the tourist destination. The second is from an economic point of view to solve tourists' destination choice decision problem among multiple destinations with microeconomic method to obtain the greatest satisfaction during the tour. The third is studied on operational research; for example, Tang Li-fan proposed the optimal tourist trails system design method based on graph theory. Domestic and foreign scholars have found that the trails system organizations play an important role in the whole travel process and begin to explore the best mode such as Campbell mode [3], Stewart-Vogt multidestination travel mode [4], Lundgren mode [5], and Jigang and Yifang mode [2]. It can be concluded that

the recent route research mainly focuses on regional scale instead of microscale study such as scenic internal travel route design. Some route designs still stay at the conceptual level and lack maneuverability. As personalized trend is increasingly evident, every tourist has its own psychological preferences and demand so that the same travel route design cannot meet all tourists' demand. So the tour route design should be given to make the tourists get the maximum satisfactions considering psychological preferences of tourists. It is very necessary for us to obtain the tourists' satisfaction degree and analyze the relationship between the tourists' satisfaction degree and travel route choice behavior to enhance the travel behavior models.

To get the reasonable tourism organization mode, the tour routes organization model of the domestic and foreign is summarized and evaluated. Considering the influence of tourists' behavior characteristics and psychological preferences, tourist routes optimal model based on multiobjective optimization function is prompted.

## 2. Behavioral Theories

This section discusses behavioral theories that may explain the dynamics in satisfaction. How do people evaluate their satisfaction? Judgments of satisfaction are influenced by the available information and by the heuristics people use at the time of making these judgments [6–10]. They present a model showing the influence of mood and comparison processes on satisfaction evaluations. Three types of comparison processes have been discussed in the literature: comparison to self, comparison to others, and counterfactuals [11].

Comparison to self involves comparing one's present situation with one's previous situation or predicted future situation. Tourists feel different ratings of satisfaction for each attraction which involves comparing one attraction with another attraction or comparing satisfaction of this tour with that of one's previous tour experience. Perceived improvements in one's situation (e.g., better health, etc.) will lead to the increases in ratings of satisfaction, but it is limited by changing aspiration levels and adaptation effects. Comparison to others (or social comparison) is the most popular type of comparison discussed and involves comparing one's own situation to that of a comparison group. People make judgments about satisfaction degree based on whether one is better (downward comparison) or worse (upward comparison) than others. Finally, counterfactuals refer to comparisons of one's current situation with hypothetical situations that did not happen but could have happened and making judgments accordingly.

Thus, comparison processes involve reference points which are used as the basis of judgment. In prospect theory, reference points are used as the basis of evaluation of outcomes; outcomes that are better than the reference point (e.g., larger monetary value) are perceived as gains and those that are worse are perceived as losses [12]. A number of studies in the transportation field have attempted to explain route choice or mode choice using prospect theory and reference points. The effects of critical incidents on car users' predicted

satisfaction with public transport were analyzed [13, 14]. Also the influence of mood on satisfaction evaluations should be discussed. The congestion level of the attraction will affect the tourists' feeling and the attitude of the attraction. The more crowded the attraction is, the less satisfied the tourist feels.

In summary, in the context of the tour route experiment described in this paper, it may be postulated that when tourists are asked about their satisfaction with the total tour and each attraction, the information they used and the processes are different from those in operation after they have been "forced" to think about their options. After experimenting with suggested tour route applied by the tourism enterprises, they would gain new information and adjust any prior misperceptions about the tour. Therefore the tourists would be more aware of the options they have. Consequently, the measure of tour satisfaction is expected to be important.

## 3. Influence Factors of Tour Route Design

Tour space is divided into large, medium, and small scales. This paper puts forward basic principles of tour route design within small scale related to the tourists diversity of spatial behavior, whose route design is within the scenic area.

*3.1. Tourists Demand Diversification.* The diversification of tourists demand can be divided into two levels.

On the microlevel, tourists with different psychological characteristics have different travel motivations which can lead to different travel behavior. The behaviors of tourists are limited by the subjective conditions such as gender, age, national, psychological interest, ability, occupation, income, education level, social status, family structure, and residence conditions which may have different influences upon the spatial behavior of tourists [15, 16]. Different tourists have different attitudes or perceptions of the same attractions. The tourists flow volume and direction in the different attractions are closely related to the level and visibility of the destination. Positive correlation exists between the attractions visibility and tourist accommodation, which lead to the tour route utilization being extremely uneven and the Gini coefficient being as high as 0.38. So the satisfaction degree of the attractions is defined and asked about in the tourists' questionnaire.

The number and the tour order of the attractions are usually different for each tourist because of the diversification of tourists demand. And to evaluate the diversification of tourists demand, Tour Route Diversification Index is defined. The larger the Tour Route Diversification Index is, the bigger the diversification of tourists demand is.

On the macrolevel, the spatial behavior exhibited some regularity which depends on the travel destination features and traffic conditions.

*3.2. Constraints Condition of Tour Route Design.* The choices of the tour route are not only affected by the tourists' subjective view but also by the tour constraints condition such as economic constraints, physical constraints, and time constraints [16]. Economic constraints make tourist save

the transport costs or the tickets cost during the choice of tour routes. Physical constraint makes tourist walk less and the number of the attractions visited during the tour is limited. Time constraints mean each tourist has a tour time budget and hopes to visit more attractions within the time budget and get the greatest satisfaction. At the same time, for each tourist, in order to complete the experience of the attractions, there should be a basic time guarantee in each attraction. The effective residence time for each attraction of the tour route is defined and investigated. The differences of the effective residence time at different attractions are caused by the characteristics of different tourists which lead to the diversification of tour route.

From the above, it can be concluded that, except for the diversification of tourist's demand, transport and tickets cost, physical constraint, tour time budget, number of attractions, and effective residence time are the important factors which will affect tourists' tour routes choice and must be considered during the tour route design.

#### 4. Tour Route Investigation and Analysis

*4.1. Survey Design and Implementation.* The survey consisted of three parts: socioeconomic and demographic characteristics, travel behavior of tour route, and the demand and tour behavior characteristic of tourists.

In the first part, socioeconomic and demographic characteristics of tourists were interviewed such as gender, age, national, psychological interest, ability, occupation, income, education level, social status, family structure, and residence conditions which may have different influence upon the spatial behavior of tourists.

The subject of travel behavior of tour route was to get some information about the scene spot such as scenic type, expenditure, tour time, total number and spatial distribution of the attractions in scenic spots, number of attractions and effective residence time, visited order of the attraction, and so forth which can support datum sustain for tour route design.

The demand and tour behavior characteristic of tourists such as tour motivations and goals and tour time budget are collected. Finally the perception and experience of the tour such as the congestion level and satisfaction degree of the attractions and the revisit preference of the tourist are investigated.

The survey takes the mode of quiz face to face in the Dajue Temple in Beijing. The survey was conducted from April 15 2012 to May 20. A total of 60 respondents were obtained.

*4.2. Analysis on Survey Results.* This section presents descriptive findings from the experiment related to tourists' characteristics, tour route choice, perceptions and attitudes, and tour satisfaction.

Socioeconomic and demographic characteristics were summarized and analyzed. About half of the participants were male. The majority of participants were between 20 and 40 years old, with an average age of 43 years. The average size of the tourist was 3.1 and most of them are between 3 and 4. Moreover, the majority of participants had high income.

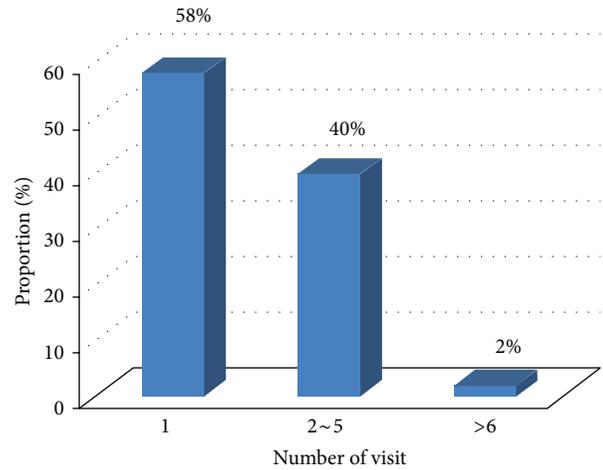


FIGURE 1: Number visit distribution.

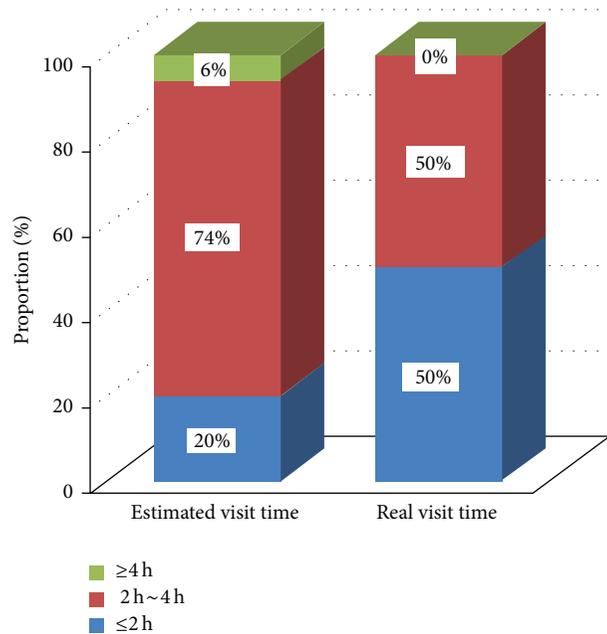


FIGURE 2: Comparison between tour budget time and real visit time.

25 percent of the participants had annual income greater than ¥100000, and 10 percent of the participants did not report their income.

As shown in Figure 1, for 58 percent of the tourists it is the first time to visit Dajue Temple and about 40 percent of the tourists revisit Dajue Temple. Trip mode is interviewed and most of the tourists came to the Dajue Temple with private car and public transit.

Each tourist has his (her) own tour time budget. It is about 74 percent of the tourists whose estimated visit time is between 2 and 4 hours as shown in Figure 2. 50 percent of the tourists finished visiting Dajue Temple within 2 hours. Comparing the tour budget time with real visiting time, it can be concluded that for about 30 percent of the tourists their real visiting time is less than their tour budget time.

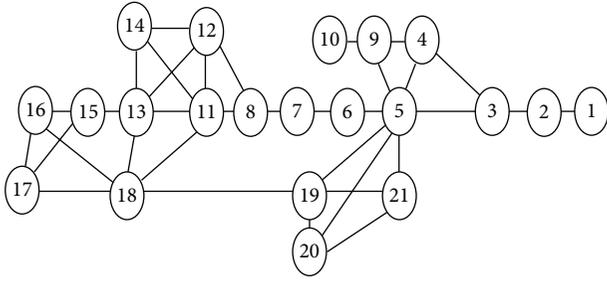


FIGURE 3: Distribution of the attractions of Dajue Temple.

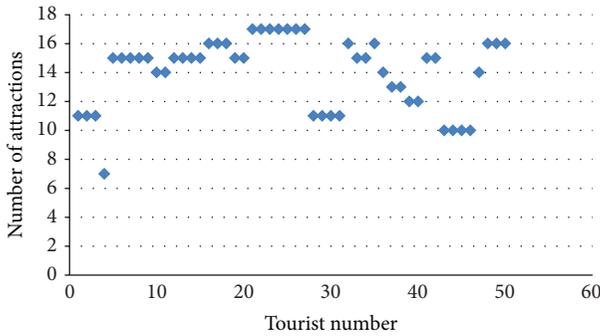


FIGURE 4: Number of attractions in tour route.

There are 21 attractions in Dajue Temple which is a famous temple and lies in the west of Beijing. The distribution of the attractions of Dajue Temple is listed in Figure 3. The total number of attractions which the tourists visited in their tour route is between 7 and 17 as shown in Figure 4. The total numbers of visited attractions by revisit tourist are less than the tourist for the first time. They will gain more satisfactions during the tour. Hence, during the tour route design, the tourist type should be considered. Numbers 6 and 3 as listed in Figure 3 are the famous attractions and all the tourists like to visit them. For some reasons such as the out-of-the-way of the attraction of the lack of traffic guilds some attractions are less visited.

Effective residence times of each attraction are gained. It can be given that most of the effective residence is between 5 and 20 minutes except 2 attractions and the average effective residence time is about 15 minutes as shown in Figure 5.

The main motivations of the tourists to Dajue Temple interviewed are as shown in Figure 6: “accompany the family, and friends” and “enjoy the scenery,” accounted for 32% and 30%, respectively; the purpose of physical exercise and relaxation accounted for 13% and 12%, respectively; religious beliefs purposes accounted for only 8%.

Measures of satisfaction with the attractions were obtained. Prior to the survey, participants rated their satisfaction on a 5-point scale anchored by “very dissatisfied” to “very satisfied,” as a response to the following question: “taking all things together, how satisfied are you with the attraction?” The satisfaction degree of the attraction is defined to be between 1 and 5. The higher the satisfaction degree of the attraction is, the better the tourists feel.

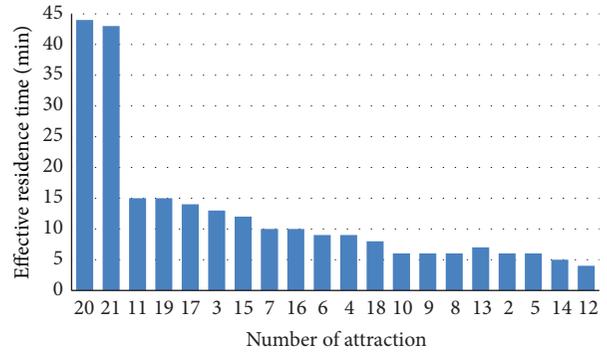


FIGURE 5: Effective residence time.

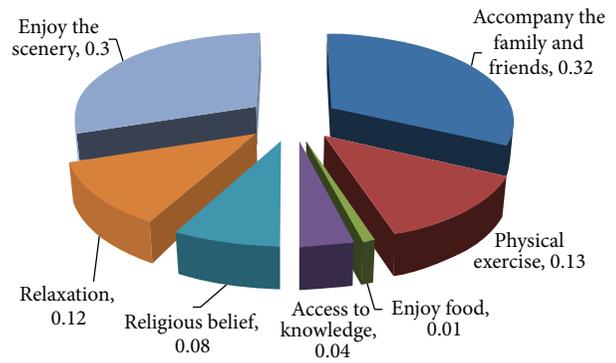


FIGURE 6: Motivation of the tourists.

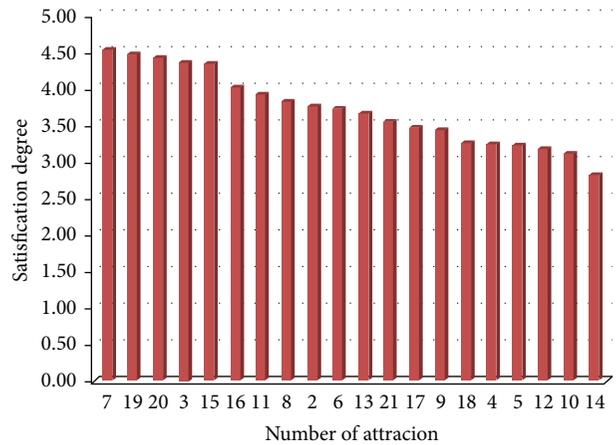


FIGURE 7: Satisfaction degree of attractions.

It can be concluded that the highest satisfaction degree of the attraction is Number 7 attraction which is 4.54 as shown in Figure 7. The average satisfaction degree of the attraction is between 3 and 5.

The relationship between the visit frequency of attraction and satisfaction degree and effective residence time and satisfaction degree are studied and listed in Figures 8 and 9. The results show that visit frequency and effective residence time and tourists attractions satisfaction are positively correlated with satisfaction degree of the attraction. The higher

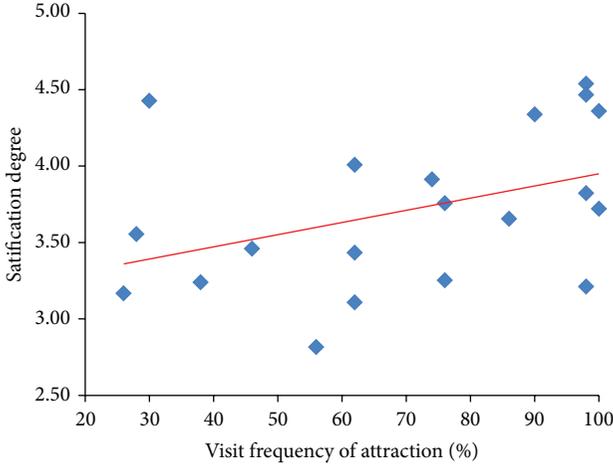


FIGURE 8: Visit frequency of attraction and satisfaction degree.

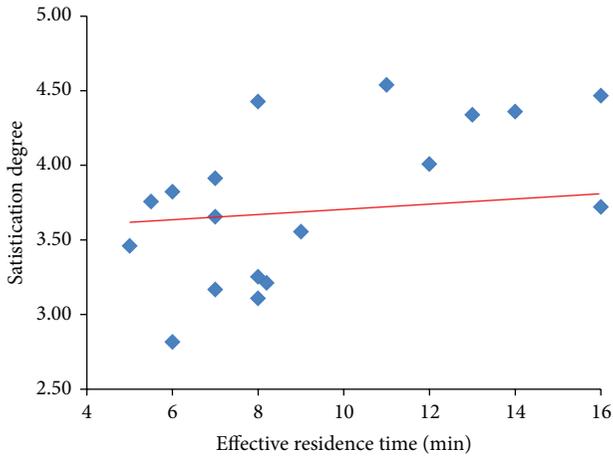


FIGURE 9: Effective residence time and satisfaction degree.

the satisfaction degree is, the more effective residence time is, and the higher the visited frequency of the attraction is.

## 5. Tour Route Design Based on Multiobjective Optimization Model

**5.1. Multiobjective Optimization Model.** During the tour, each tourist has a tour time budget and the tourists hope to visit more attractions within the time budget and get the greatest satisfaction. It can be concluded that optimal design of tour route is a multiobjective optimization procedure. The multiobjective optimization model is carried out for the tour route design.

Multiobjective optimization problem is listed as follows [17]:

$$\min y = f(x) = \{f_1(x), f_2(x), \dots, f_n(x)\},$$

$$n = 1, 2, \dots, N,$$

$$\begin{aligned} \text{s.t. } & g_i(x) \leq 0, \quad i = 1, 2, \dots, m, \\ & h_j(x) = 0, \quad j = 1, 2, \dots, k, \\ & x = \{x_1, x_2, \dots, x_D\}, \end{aligned} \quad (1)$$

where  $x$  is the decision vector of  $D$ ,  $y$  is objective vector,  $g_i(x)$  is the  $i$ th inequality constraints,  $h_j(x)$  is the  $j$ th equality constraints, and  $f_n(x)$  is objective function.

Two indicators are considered during the optimal design of the tourist routes. Optimized objective functions are described as follows:

$$\begin{aligned} \text{Min } & \sum_n I_E^i \cdot x_i, \\ \text{Max } & \sum_n I_T^i \cdot x_i, \end{aligned} \quad (2)$$

where  $x_i = 1$  which means the  $i$ th attraction is on the tour route and  $x_i = 0$  which means the  $i$ th attraction is not on the tour.  $I_E^i$  is the distance between  $i-1$ th and  $i$ th attraction;  $I_T^i$  is the satisfaction degree of the  $i$ th attraction.

Because the tourists are more concerned with the satisfaction degree of the attraction during the optimal design of the tourist routes than the distance, the maximum satisfaction degree of the attraction is set as the first optimization objective and the minimum distance is set as the second optimization objective. The multiobjective optimization model is expressed as shown in

$$\min z = P_1 (d_1^-) + P_2 (d_2^+), \quad (3)$$

$$\text{s.t. } \sum_i I_T^i \cdot x_i + d_1^- + d_1^+ = E_{\max}, \quad (4)$$

$$\sum_i I_E^i \cdot x_i + d_2^- + d_2^+ = E_{\min}. \quad (5)$$

In addition, according to the tourists demand, the constraint of the tour time and number of the visited attractions are listed as follows:

$$0 < T \leq T_{\max}, \quad (6)$$

$$i_{\min} \leq i \leq i_{\max},$$

where  $P_1$  and  $P_2$  are priority factors,  $P_1$  is first priority factor,  $P_2$  is second priority factor,  $E_{\max}$  is the maximum satisfaction degree,  $d_1^-$  is the negative deviation value degree,  $d_1^+$  is the positive deviation value,  $E_{\min}$  is the minimum length of tour route,  $d_2^-$  is the negative deviation value, and  $d_2^+$  is the positive deviation value.

**5.2. Calculation of the Multiobjective Optimization Model.** The level algorithm is carried out to calculate the multiobjective optimization model.

Step 1: optimizing the  $P_1$  level in objective function

$$\max z_1 = d_1^-. \quad (7)$$

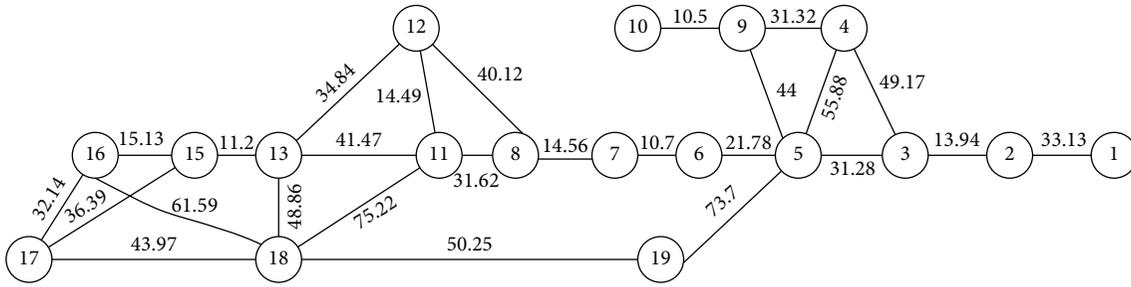


FIGURE 10: Distance.

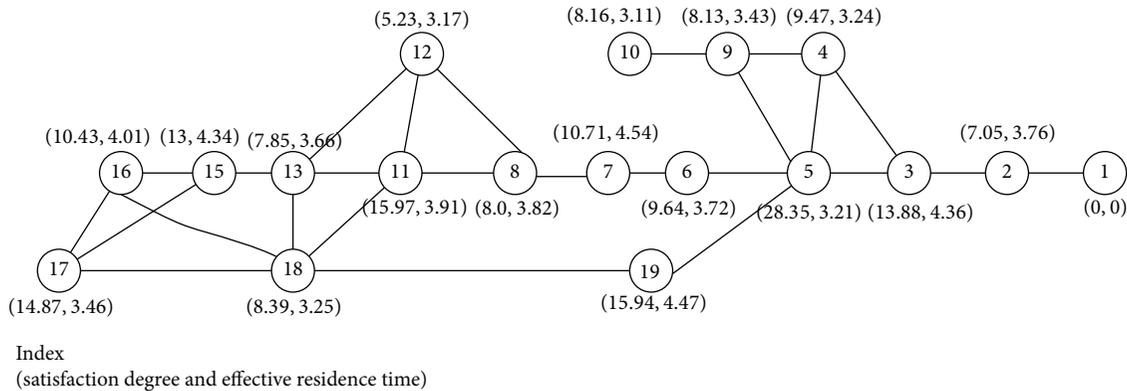


FIGURE 11: Effective residence time and satisfaction degree.

TABLE 1: Tour time.

Tour time	Proportion (%)
≤3	45
>3	55
Total	6

Step 2: optimizing the  $P_2$  level in objective function

$$\min z_2 = d_2^+. \tag{8}$$

Step 3: Analysis of deviation: deviation index of satisfaction degree ( $\sigma_s$ ) and tour route distance ( $\sigma_d$ ) are defined and calculated.

Step 4: Based on the particle swarm optimization model, the priorities of the tour routes are calculated [17–19].

5.3. *Sample.* The distance and effective residence time are listed in Figures 10 and 11.

According to the questionnaire about tour route tourist, the number of visited attractions is between 7 and 17. The available routes are listed and categorized by the tour time as listed in Table 1.

It can be seen that 45 percent of the tourists' tour time is less than 3 times and 55% is larger than 3 hours. So the route which tour time is less than 3 hours is defined as quick tour and the route which tour time is larger than 3 hours is defined as depth tour.

The deviations of two routes are calculated, respectively, and listed in Table 2 and Table 3.

Deviation index of satisfaction degree ( $\sigma_s$ ) and tour route distance ( $\sigma_d$ ) are calculated. Deviation index of satisfaction degree is listed in Table 3. For the depth tour route, only when  $\sigma_s$  is between 1 and 0.18 and  $\sigma_d$  is between 0 and 0.104, the route is the acceptable result. For the quick tour route, only when  $\sigma_s$  is between 0 and 0.053 and  $\sigma_d$  is between 0 and 0.172, the route is the acceptable result.

The priority of the tour route is calculated and the suggested depth tour route and quick route are given, respectively, as shown in Figure 12.

Comparing the depth tour route with quick route, there are some differences among total number of the attraction, tour order, effective residence time, and so forth. For the suggested depth route, the total number is less than the quick tour route. Some attractions which are less famous are introduced so that the revisited tourists can experience and enjoy the tour route.

## 6. Conclusion

Conclusions can be given as follows.

A questionnaire about tourists' tour route and satisfaction was carried out and some information about the scene spot and tourists demand and tour behavior characteristic is concluded and analyzed. Visit frequency and effective residence time are positively correlated with satisfaction degree of the attraction. Comparing the tour budget time with real visiting

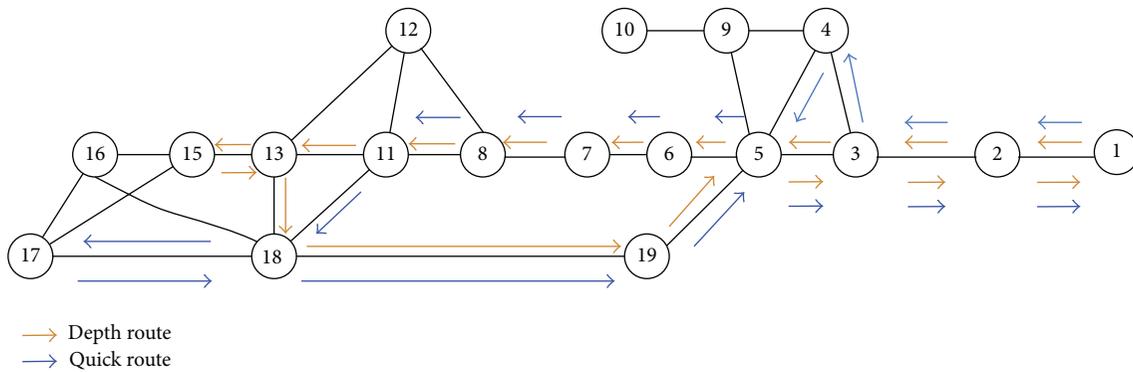


FIGURE 12: Suggested depth tour route and quick route.

TABLE 2: Parameter of depth tour and quick tour.

	Satisfaction degree		Distance		Tour time	
	Depth tour	Quick tour	Depth tour	Quick tour	Depth tour	Quick tour
Max.	3.40	3.44	672.73	670.68	3.61	2.99
Min.	2.99	2.69	488.70	388.99	3.01	1.72

TABLE 3: Deviation index of satisfaction degree.

$\eta$	$\sigma_s$		$\sigma_d$	
	Depth tour	Quick tour	Depth tour	Quick tour
	0.018	0.053	0.104	0.172

time, it can be concluded that for about 30 percent of the tourists their real tour visiting time is less than their tour budget time.

Based on the convey datum, tour routes multiobjective optimization functions are prompted considering the tourists' behavior characteristics and psychological preferences. According to the distribution of the route time, the route which tour time is less than 3 hours is defined as quick tour and the route which tour time is larger than 3 hours is defined as depth tour. The priorities of the tour routes are calculated and the suggested depth tour route and quick route are given, respectively. The results show that visit frequency and effective residence time and tourists attractions satisfaction are positively correlated with satisfaction degree of the attraction.

Furthermore, during the tour route convey, it can be found that the congestion level of the attractions will affect the tourists' satisfaction degree. The more crowded the attraction is, the less satisfied the tourist feels. Also tourists will reduce the effective residence time in the crowded attractions and change their tour route under the information provided and distributed by the tourism enterprises. So the mechanism and influence of the congestion level on the tour route should be studied in the future.

### Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

### Acknowledgments

This research was supported by National Basic Research Program of China (no. 2012CB725403) and the National Natural Science Foundation of China (Grant no. 51308015). The authors are very grateful for the comments from the anonymous reviewers. The authors would like to thank one anonymous referee for his (her) helpful comments and suggestion, which improved the content and composition substantially.

### References

- [1] R. A. Abbaspour and F. Samadzadegan, "Time-dependent personal tour planning and scheduling in metropolises," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12439–12452, 2011.
- [2] B. Jigang and C. Yifang, *Geography of Tourism*, vol. 42, Higher Education Press, Beijing, China, 1999.
- [3] C. K. Campbell, *An Approach to Research in Recreational Geography*, Department of Geography, University of British Columbia, British Columbia, Canada, 1967.
- [4] S. I. Stewart and C. A. Vogt, "Multi-destination trip patterns," *Annals of Tourism Research*, vol. 24, no. 2, pp. 458–461, 1997.
- [5] J. Lundgren, "The development of tourist travel system: a metropolitan economic hegemony par excellence," *Jahrbuch für Fremdenverkehr*, pp. 62–65, 1972.
- [6] R. L. Oliver, "A cognitive model of the antecedents and consequences of satisfaction decisions," *Journal of Marketing Research*, vol. 17, no. 4, pp. 460–469, 1980.
- [7] L. R. Allen, H. R. Hafer, and P. T. Long, "Rural residents' attitudes toward recreation and tourism development," *Journal of Travel Research*, vol. 31, no. 4, pp. 27–33, 1993.
- [8] P. T. Long, R. R. Perdue, and L. R. Allen, "Rural resident tourism perceptions and attitudes by community level of tourism," *Journal of Travel Research*, vol. 28, no. 3, pp. 3–9, 1990.
- [9] S. J. Calver and S. J. Page, "Enlightened hedonism: exploring the relationship of service value, visitor knowledge and interest, to

- visitor enjoyment at heritage attractions,” *Tourism Management*, vol. 39, pp. 23–36, 2013.
- [10] J. L. McElroy, P. Tarlow, and K. Carlisle, “Tourist harassment: review of the literature and destination responses,” *International Journal of Culture, Tourism and Hospitality Research*, vol. 1, no. 4, pp. 305–314, 2007.
- [11] M. Abou-Zeid and M. Ben-Akiva, “The effect of social comparisons on commute well-being,” *Transportation Research A*, vol. 45, no. 4, pp. 345–361, 2011.
- [12] D. Kahneman and A. Tversky, “Prospect theory: an analysis of decision under risk,” *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979.
- [13] E. Avineri and P. H. L. Bovy, “Identification of parameters for a prospect theory model for travel choice analysis,” *Transportation Research Record*, vol. 2082, pp. 141–147, 2008.
- [14] T. Pedersen, P. Kristensson, and M. Friman, “The focusing illusion in travel mode preference: effects of critical incidents on car users’ predicted satisfaction with public transport,” in *Proceedings of the 12th International Conference on Travel Behavior Research*, Jaipur, India, December 2009.
- [15] R. Prentice, “Community-driven tourism planning and residents’ preferences,” *Tourism Management*, vol. 14, no. 3, pp. 218–227, 1993.
- [16] H. Yan and G. Hongzhi, “Study on models of commuter mode choice beyond fuel prices based on ordered logit models,” *Journal of American Science*, vol. 6, no. 8, pp. 230–235, 2010.
- [17] H. F. Lewis and T. R. Sexton, “On the tour partitioning heuristic for the unit demand capacitated vehicle routing problem,” *Operations Research Letters*, vol. 35, no. 3, pp. 374–378, 2007.
- [18] W. Souffriau, P. Vansteenwegen, J. Vertommen, G. V. Berghe, and D. V. Oudheusden, “A personalized tourist trip design algorithm for mobile tourist guides,” *Applied Artificial Intelligence*, vol. 22, no. 10, pp. 964–985, 2008.
- [19] M. Gong, L. Jiao, D. Yang, and W.-P. Ma, “Research on evolutionary multi-objective optimization algorithms,” *Journal of Software*, vol. 20, no. 2, pp. 271–289, 2009.

## Research Article

# The Research of the Driver Attention Field Modeling

**Pengfei Tao,<sup>1,2</sup> Hongyu Hu,<sup>2</sup> Zhenhai Gao,<sup>1</sup> Xin Liu,<sup>3,4</sup> Xianmin Song,<sup>2</sup> Yan Xing,<sup>2</sup> Yuzhou Duan,<sup>2</sup> and Fulu Wei<sup>2</sup>**

<sup>1</sup> State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130022, China

<sup>2</sup> College of Transportation, Jilin University, Changchun 130022, China

<sup>3</sup> School of Management, Jilin University, Changchun 130022, China

<sup>4</sup> Network Center, Jilin University, Changchun 130022, China

Correspondence should be addressed to Hongyu Hu; huhongyu@jlu.edu.cn

Received 28 September 2013; Accepted 4 December 2013; Published 6 January 2014

Academic Editor: Huimin Niu

Copyright © 2014 Pengfei Tao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For expanding the application scope of car-following, based on the basic idea of the noncontact interaction of the objects in physics, establish an attention field model to describe the driving behavior. Firstly, propose the time distance concept to describe the degree of driver perception to the front one-dimensional space and extend its application range to the two-dimensional space. Secondly, connect the point which has the same time distance to constitute the equipotential line of drivers' attention field equipotent, and establish a model to describe it. Thirdly, define the effective range of the driver's psychological field with the feature of the driver's visual distance range increasing and the angle decreasing. Finally, design the calculation method to collect projection of the object in the psychological field scope and calculate the curve points to determine the object's intensity of psychological field. Preliminarily build the driving behavior model and use the numerical simulation method to simulate the simple transport scenarios; initially verify the validity of the model.

## 1. Introduction

The research on the driving behavior has been the focus in the traffic engineering study. The driving behaviors model establishes a link between the characteristics of the single vehicle and the traffic. It explains the macrotransport phenomena from the moving of the microindividual and makes the development of the traffic flow theory. The driving behaviors model has important applications in the field of traffic simulation, traffic safety, and intelligent driving.

The car-following model is the main form of the model driving behavior. Car-following model was used to describe the driving behavior of following vehicle facing the actions of the leading vehicle in the single lane with limit of overtake. It played an important role in the traffic theory and was widely used in the fields of microscopic traffic simulation, smart driving, traffic safety, and so forth. The researches of car-following have been lasting for years and made a series of results [1–3], and the GM car-following model

was proposed firstly and considered the most classic model in those research results [4, 5]. This model regarded the difference between the leading and following car as the source of stimulation and combined with the space headway and velocity to obtain the driver's acceleration decision making [6–9]. A variety of car model were from derived; the GM modeling idea [10–13]. Lee proposed a GM model with memory effect; the basic idea was that the driver react relied on the stimulation of a period of time rather than the point-in-time. Herman established a model to describe the impact on driver form the multivehicle in the front; it considered that the driver received the effect not only from the single leading car but also from the continuous traffic flow before the him.

With the developing of microscopic traffic simulation technology, the microdriving behavior of single vehicle is focused on [14, 15]. The models become complex more and more and tend to descript the specific behavior for different conditions to make the simulation realistic. And basing on it, obtain the macrotraffic situation by simulating a number

of running individual vehicles. The typical achievements contain mental-physical model, fuzzy inference model, safe distance model, and so forth.

A fundamental assumption of the traditional car-following model is that all vehicles are travelling in the center of the lane. But it is difficult to achieve this ideal state. Gunay (2007) studied the transverse distribution of the vehicle in the lane, it showed the vehicle occurrence frequency decrease from the center to sides in the lane, similar to normal distribution, and exhibited the lateral offset phenomenon which was prevalent. Meanwhile, the traditional car-following model only considered the effect from the leading vehicle but ignored the front vehicle in the adjacent lane, and this would exert a certain extent impact on the driver.

In addition, most traditional car-following models research the model simply based on the movement of the vehicle. The driver is the operator of the vehicle; his perception of traffic environment would determine the state of the vehicle motion. The previous researches mostly focus on the rules of vehicle running trajectory or the certain physiological thresholds as the trigger of driving operation. Relatively lack the considering of driver's characteristic.

This paper proposes a driver attention field for describing the driver's attention distribution, and establishes a driving behavior model framework with considering the drivers' individual characteristics.

## 2. Method

*2.1. The Basic Idea of Modeling.* Driving behavior is the combined effect of the driving task and the driving environment to the drivers' psychological. The driving task encourages the driver to drive toward the goal direction with the desired speed and leads to a promote effect for the driver's psychological. This impact would change with the characteristics and the motion of the drivers. For example, if the velocity is low, the psychological impact would be much more than normal and force the driver to speed up. The psychological impact is also related with the driving states.

And the driving environment may prevent the driver from selecting a higher speed and the corresponding movement direction, causing a psychological stress for driver. This psychological stress reflects the driver's different levels attention to the objects in driving environment. It is related to motion state, attribute and position of the object. For example, the driver would pay more attention to the objects in front than lateral. Similarly, the driver would take notice of the objects with high speed and short distance.

Driving behavior is the balance process under the combined effect of psychological driving and pressure. When the psychological drive has more effects than the psychological pressure, the driver would choose accelerating to reduce the psychological drive. On the contrary, the driver would decelerate or steer the driving direction to reduce the psychological stress. The driver's behavior is running under this balancing process.

This paper takes the distribution of driver's attention in space as a field and applies the field theory to describe the

behaviors of the driver response to the surrounding vehicles, obstacles, signs and markings, and so forth. This attention field is different from the field in physical. The generating source of field in physical would influence the objects in field. The generating source of driver attention field is the driver, and any objects in the field would not be impacted. But the driver would react against the environment under the constraints of its own characteristics. It means the driver determines the strength of the impacts from the objects in driving environment and changes his driving velocity or direction to avoid or reduce the impact of the environment.

In this paper, we design the driver's attention field as a scalar field and reflect the driver's perception of the surrounding traffic environment with the changing of the field intensity value. The field intensity is the function of the field source (driver) relative position; it is only related to the driver's own characteristics and the motion state of his vehicle. For the objects in the driving environment, just as the positive and negative charged particles in an electric field would be completely different effects, the different types and motion would be differently impacted. It needs to be modeled according to the specific circumstances.

There are a number of similar natures between the attention field and physical field, such as the ultradistance effect; the impact results are not only due to the field itself, but also depend on the properties of objects.

But both principle and the nature of the two fields exist essential difference. In the physical field, such as the electrical fields, the objects in the field would be generated by the source of ultra-distance effect, causing the motion state transform. The objects in the field would be generated by the source of ultradistance effect, causing the motion state transform. The attention field exists in the driver's inner perception. It is not able to affect the objects in it and change the states of objects. But the field could transmit the psychological pressure or driving force to the driver. The driver adjusts his own motion (such as speed up or down, change the direction) to mitigate the pressure impacts from objects and achieves the ultra-distance effects.

*2.2. Establish the Modeling Frame.* Analyse the impact of the driver from the driving environment, as Figure 1 shows. The vehicle  $n$  is the study object, which exists around an attention field; we also call it target vehicle below. And set the vehicle locate in the field range  $n - 1$ .

The attention field of vehicle  $n$  reflects the time the vehicle  $n$  reaches a different spatial locations and also shows the perception of vehicle  $n$  driver about the threat degree from the object in this locate.

The attention field of vehicle  $n$  driver is created by the driver's subjective perspective. The vehicle  $n$  driver could not accurately determine the  $n - 1$  car driver's next move. So the vehicle  $n$  driver would adjust his operating status (including velocity and direction) to achieve the ultra-distance effect to himself. The magnitude and direction of the effect are closely related to the attention field intensity of vehicle  $n - 1$  location and the relative speed between the vehicle  $n$  and  $n - 1$ .

It could be more general; expand the vehicle  $n - 1$  to arbitrary objects, including stationary obstacles, traffic signs,

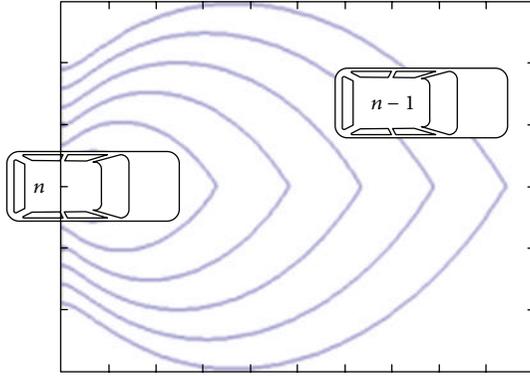


FIGURE 1: The attention field schematic.

and so forth. The impact of vehicle  $n$  driver could be described by the attention field proposed above.

The basic idea of modeling the attention field is as follow: firstly, base on the judgment of the field distribution in the space, secondly, analyse the distribution characteristics of equipotential lines, and finally, build a specific potential field values to get the field intensity distribution model.

The attention field exists in the target vehicle moving plane. There must be many equipotential lines in it. If the form of the equipotential lines model is ascertained, the specific calculation method could be relatively easily gotten.

In the vehicle traveling direction of a one-dimensional line, the reciprocal of the time distance could express the form of the field intensity function. The time distance refers to the time of the vehicle reaches the specified position with current speed. Obviously, the greater the time distance is, the lower the attention the driver pays to the front position, which means the lower the field intensity.

Obviously, in case of a steady velocity of target vehicle and a unchanging angle between the target vehicle travelling direction and point-target vehicle line, the closer to the target vehicle, the greater attention field intensity is. And in the case of a fixed distance, the bigger the angle between the direction of the target car driving and the point-target vehicle lines, the weaker the field strength is.

As shown in Figure 2, set the point  $P_1$  along the direction the target vehicle travel direction. The curve contains the point  $P_1$  and  $Q_1$  is an arc of a circle with the center point  $O$ . It is clear that with the same distance to point  $O$ , the line  $OQ_1$  has the bigger angle with target traveling direction than  $OP_1$ . According to the previous relationship between field intensity and angle, the field intensity at point  $P_1$  is stronger than  $Q_1$ . In the same angle, the closer to the target vehicle the bigger the field strength, and there is a point with the same field intensity of  $P_1$  between the point is  $O$  and  $Q_1$ ; let it be  $P_2$ . Similarly, deduce  $P_3, P_4$ , and many other potential points with the field intensity equal to  $P_2$ . Connect these points and obtain the approximate shape of equipotential lines as Figure 2 shows.

It can be considered that the driver's attention field intensity is the function of the time-distance in the traveling direction.

When close to the target vehicle, the small time distance would produce relatively large psychological pressure

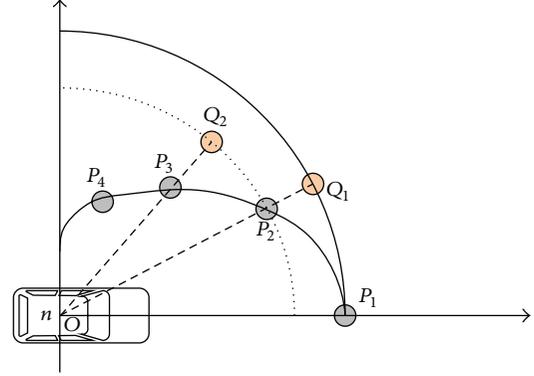


FIGURE 2: The attention field equipotential line shape schematic.

and from a big field intensity. And in the position away from the target vehicle, the driver would reduce his concern degree with the ample response time, and the field intensity is relatively small.

Thus, determine that the field intensity basic form in the traveling direction is the function of the reciprocal of the time-distance as follows:

$$Eb_n(P_x) = \frac{v_n + v_\epsilon}{L_0} \tag{1}$$

The  $P_x$  is the point in the target vehicle traveling direction. The  $Eb_n$  is the basic field intensity; there is positive correlation with the field intensity.  $v_n$  is the velocity of the target vehicle driver in.  $L_0$  is the distance of between the target and the point  $P_x$ .  $v_\epsilon$  is the correction value.

2.2.1. Establish the Equipotential Lines. Because the field intensity value could be determined by the target vehicle speed and the distance in the traveling direction, choosing a particular point on the traveling line as the reference point and constructing the corresponding equipotential lines according to that point are needed.

Obviously, the shape of the equipotential lines is related to the driver's individual character. The cautious drivers would invest more attention to the space on the sides than aggressive ones. And in addition, the equipotential lines shape also depends on the velocity of target vehicle. In the case of the high speed, the driver would reduce his attention to the sides space, and the shape of the equipotential lines would become flat, and if the vehicle is running with the low speed, compared with the previous case, the equipotential lines shape would be relatively mellow.

For the point in the traveling direction of the target vehicle, after determining the field intensity value of this point, it needs to find the set of points with the same intensity value in the space. Then model according to the general form equipotential lines.

The point in equipotential line with the specific field intensity value would get near the target vehicle with the angle increasing, and when angle is zero, the distance between the target vehicle and the point is biggest. If the point is at the vertical direction of the vehicle, the distance between

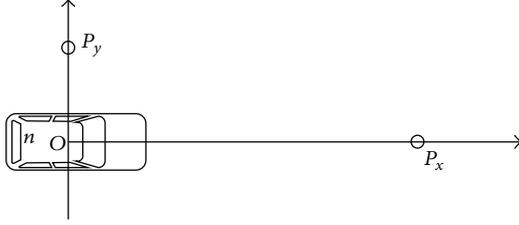


FIGURE 3: The two special points in the attention field equipotential line.

the target vehicle and the point reach is the smallest. There should be a link between the minimum and maximum values.

In Figure 3, the point  $P_y$  is in the direction vertical to the vehicle travel. The point  $P_x$  is on the vehicle traveling direction.

If the points are fixed, with the speed of target vehicle  $n$  increasing, the time-distance in point  $P_x$  would reduce, and the attention field intensity negative correlation with it would increase.

Based on the general driving behavior, the driver's concern degree has a more significant reduction from the front space to besides. And with the high speed, the driver's vision field width would be smaller, and pays less attention to the besides space. So in the fixed point  $P_y$ , the corresponding field intensity value would decrease with the target vehicle increasing.

Assume the two points  $P_x$  and  $P_y$  are on the same equipotential line. After determining the field intensity value of the equipotential line, as the velocity increasing, the point  $P_x$  in this equipotential line would move forward, and the point  $P_y$  would move toward the target vehicle, from the shape as Figure 4 shows. Obviously, with the speed getting big, the equipotential line shape would get sharp; if the speed is low, the equipotential line shape would get relatively round.

Meanwhile, like the field physics, the equipotential lines in attention field should have a nonoverlapping feature. It could assume that the ration between the  $OP_x$  and  $OP_y$  is fixed.  $OP_x$  is the distance from intersection between the target vehicle driver's attention equipotential line and the vehicle traveling direction to vehicle  $n$ .  $OP_y$  is the distance from the vehicle  $n$  position to the intersection between the attention equipotential line and the direction vertical the vehicle traveling. Let the ratio be  $\alpha$ . The  $\alpha$  is associated with the target vehicle speed, and also associated with the driver's own property. Clearly, based on the above analysis and Figure 4, for the same type of driver, the higher the speed, the smaller the  $\alpha$ .

The drivers could be divided into three types: cautious, aggressive, and balanced. Different types of drivers have different perceptions of the surrounding environment. It is embodied in the different-value  $v_\varepsilon$  in formula (1). The cautious driver correspond to a relatively bigger  $v_\varepsilon$ , the aggressive type correspond to a smaller  $v_\varepsilon$ , and the  $v_\varepsilon$  value of balanced driver is in the middle.

In Figure 4, the angle  $\theta$  of the equipotential line in the first quadrant of the coordinate system with the vehicle  $n$  as origin increases with point  $P$  getting close to the target vehicle  $n$  and becomes max at point  $P_x$ . The field intensity of

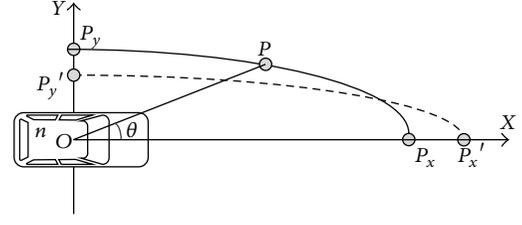


FIGURE 4: Attention field equipotential line trend schematic.

this equipotential line is determined by the velocity of vehicle  $n$  and the length of  $OP_x$ . Let the length of  $OP_x$  be  $L_0$ . And the mathematical expression of the corresponding equipotential line is shown as

$$\rho = L_0 - (1 - \alpha) L_0 |\sin \theta|. \quad (2)$$

The  $P$  is an arbitrary point in equipotential line. Point  $O$  is the position of vehicle  $n$ . The  $\rho$  is distance between the point  $P$  and the vehicle  $n$ .  $\theta$  is the angle between the  $OP$  and the direction (the  $x$ -axis in Figure 4) of vehicle  $n$  traveling.  $\alpha$  is the ratio between  $OP_x$  and  $OP_y$ ; its value is related to the speed of vehicle  $n$ . The  $L_0$  is the distance between the target and the point  $P_x$ . It is related to the speed of vehicle  $n$  and the corresponding field intensity of equipotential line.

2.2.2. *The Field Intensity Modeling.* It can obtain the formula (3) from (2):

$$L_0 = \frac{\rho}{1 - (1 - \alpha) |\sin \theta|}. \quad (3)$$

For any point  $(x, y)$  in the space like Figure 4, the length  $\rho_{(x,y)}$  between the origin (vehicle  $n$  position) and it can be easily obtained. The parameter  $\alpha$  is only related to the target vehicle speed. So the parameter  $L_0$  could be obtained. Then combined with the formula (1), calculate the field intensity in point  $(x, y)$ . Finally, obtain the method to calculate the attention field intensity of arbitrary point in the space, as follows:

$$Eb_n(x, y) = \frac{v_n + v_\varepsilon}{\rho_{(x,y)}} \left[ 1 - (1 - \alpha(v_n)) |\sin \theta_{(x,y)}| \right]. \quad (4)$$

2.2.3. *Model frame.* The driving behavior model should include two parts: one is the driver's attention field description, and the other is the attribute information and motion state of objects in the field. Similar to the physical field, the driving behavior based on the attention field model framework follows:

$$F_n = Eb_n \cdot g_{n-1}(v_{n-1}, v_n). \quad (5)$$

The  $n - 1$  is vehicle in the target vehicle  $n$  driver's attention field is as in Figure 1.  $v_{n-1}$  and  $v_n$  are the velocity of vehicles  $n$  and  $n - 1$ .  $g_{n-1}$  is the characteristics of vehicle  $n - 1$  itself; it is relative to the relative speed and direction.  $F_n$  is the effect of vehicle  $n$ .

Form the visual point of view, the changing of  $n - 1$  vehicle position should affect the vehicle  $n$  driver, but it is not directly reflected in the model frame. This is because this fact is fully taken into account in calculating the attention field intensity.

The  $F_n$  in the modeling process usually consider setting it as the acceleration of target vehicle. And it is closely connected to the force affect the object in physical can directly reflect the attention effects.

### 3. Model Analysis

**3.1. Parameters Analysis.** In the attention field, the  $\sin\theta_{(x,y)}$  and  $\rho_{(x,y)}$  in formula (4) determine the space position;  $v_n$  is the velocity of the vehicle. The  $v_\varepsilon(\cdot)$  is used to distinguish the different characteristics of the driver's attention field and describe the effect from the vehicle movement on the driver.

$v_\varepsilon(\cdot)$  is affected by two factors: one is the driver's personality characteristics, the cautious driver corresponding to relatively big  $v_\varepsilon$ , and the aggressive driver corresponding to small  $v_\varepsilon$ . And the other factor is changing of the vehicle movement state. When the vehicle is in the accelerated state, the focus on the objects would be strengthened. It is shown as the field intensity of same position in attention field would increase, and the  $v_\varepsilon$  becomes bigger. If the vehicle is decelerating, the attention to the front would reduce, and the  $v_\varepsilon$  becomes smaller. In these two factors, the driver's personality characteristics should play a major role, while the status of vehicle is running relative to a supporting role.

The  $v_\varepsilon$  is shown as formula (6); the  $\varphi_1$  and  $\varphi_2$  undetermined coefficient.  $D_n$  is the membership function of driver personality characteristics. Its ranging is  $D_n \in [0, 1]$ ,  $D_n$  is increasing with the driver's personality aggressiveness increasing. When  $D_n(\text{driver}) = 1$ , it means the driver is extreme caution; if  $D_n(\text{driver}) = 0$  it means the driver is extremely reckless:

$$v_\varepsilon = \varphi_1 e^{D_n(\text{driver})} + \varphi_2 (e^{A_n(a_n, a_{\text{acc. max}}, a_{\text{dec. max}})} - 1). \quad (6)$$

$A_n$  is the ratio between the current acceleration and the maximum acceleration; it is shown as formula (7). The  $a_n$  is the current acceleration of the vehicle; it is a sign scalar.  $a_{\text{acc. max}}$  and  $a_{\text{dec. max}}$  are the maximum acceleration and deceleration; they are the positive scalar:

$$A_n(a_n, a_{\text{acc. max}}, a_{\text{dec. max}}) = \begin{cases} \frac{a_n}{a_{\text{acc. max}}} & a_n > 0, \\ 0 & a_n = 0, \\ \frac{a_n}{a_{\text{dec. max}}} & a_n < 0. \end{cases} \quad (7)$$

The range of  $A_n$  is  $A_n \in [-1, 1]$ . By the characteristics of the exponential function,  $e^{D_n}$  and  $e^{A_n} - 1$  are all the monotonically increasing functions as Figure 5 shows. When the character gets close to reckless driver, the value of  $e^{D_n}$  is increasing, and the changing amplitude is also increased. It corresponds to the aggressive driver having a low degree concern for the obstructions, and the cautious driver is tense to deal with the objects in front.

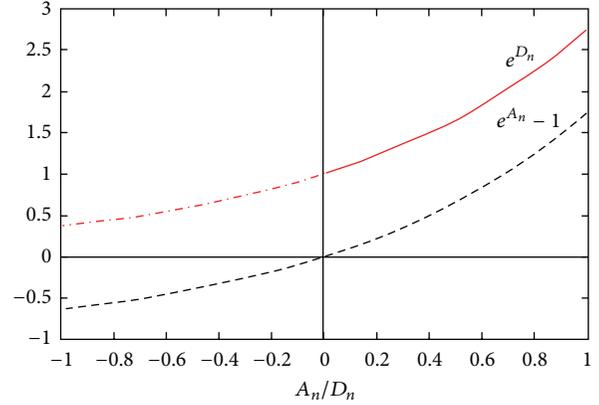


FIGURE 5: The value changing of  $e^{D_n}$  and  $e^{A_n} - 1$ .

The  $\alpha(v_n)$ , in formula (4), its value increases monotonically with decreasing speed.  $\alpha(v_n)$  is defined as the formula (8). The  $v_{\text{max}}$  is the maximum speed the vehicle could reach. The  $\alpha_{\text{max}}$  is the value of  $\alpha$  when the vehicle is stopping. And  $\alpha_{\text{min}}$  is the value of  $\alpha$  when the vehicle reaches the maximum speed:

$$\alpha(v_n) = \frac{\alpha_{\text{min}} - \alpha_{\text{max}}}{v_{\text{max}}} v_n + \alpha_{\text{max}}. \quad (8)$$

The final attention field intensity is as follows; the  $\gamma_1$  is adjustment factor:

$$E_n = \gamma_1 \cdot E b_n. \quad (9)$$

### 3.2. The Improved Model

**3.2.1. The Improved Field Intensity of Objects.** The model previously established can describe the attention field distribution. Because there are size differences of the objects in field, the field intensity of objects space is hard to represent with field strength at a point. It considers using the first curve integral to solve this problem.

The first form curvilinear integral is a commonly math method. Let  $L$  be a smooth curve in plane  $xOy$ . The function  $f(x, y)$  is bounded in  $L$ . The integral of the function on curve  $L$  is  $\int_L f(x, y) ds$ .

For the driver of vehicle  $n$  in Figure 6, limited by the vision, without considering the object height, in Figure 6, the  $n$  vehicle driver's visual range of vehicle  $n - 1$  are  $Q_1Q_2$  and  $Q_2Q_3$ . So we could simplify the  $n - 1$  vehicle with complicated shape to two segments  $Q_1Q_2$  and  $Q_2Q_3$ . Then apply the first form curvilinear integral method to calculate the attention field intensity of the vehicle  $n - 1$ .

However, the objects in the attention would be diverse and irregular, causing hard calculation of the first form curvilinear integral. Consider using the projection of the object borders in the driver's visual range to represent the object's position. Then calculate the field intensity with the first form curvilinear integral of this projection; the result is the object's field intensity, such as the vehicle  $n - 1$  in Figure 6.

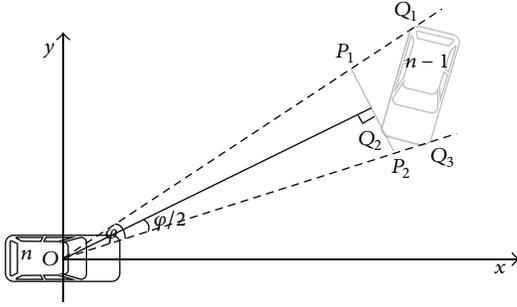


FIGURE 6: Object projection in driver's visual range schematic I.

The projection line  $P_1P_2$  is determined by angle  $\angle Q_1OQ_3$  and the nearest point  $Q_2$  to the driver on object vehicle  $n - 1$ . Let  $\angle Q_1OQ_3$  value be  $\varphi$ , and the line passing through the point  $Q_2$  and vertical to the angular bisector of  $\angle Q_1OQ_3$  has two intersection points with  $OQ_1$  and  $OQ_2$  which are  $P_1$  and  $P_2$ .

Assume the coordinate of  $P_1, P_2$  are  $(x_1, y_1)$  and  $(x_2, y_2)$ , and let  $x_1 < x_2$ . When the segment  $P_1P_2$  is not perpendicular to the traveling direction (the  $x$ -axis in Figure 6), the linear equation is as follows:

$$y = \frac{y_2 - y_1}{x_2 - x_1} (x - x_1) + y_1. \quad (10)$$

Let

$$x = t, \quad y = \frac{y_2 - y_1}{x_2 - x_1} t + \frac{y_1 x_2 - y_2 x_1}{x_2 - x_1}. \quad (11)$$

Then, the curve integral of the segment in driver's attention field is shown in formula (12). The  $E_n(x, y)$  is the field intensity model in formula (4):

$$\begin{aligned} E_n(n-1) &= \int_{P_1P_2} E_n(x, y) ds \\ &= \int_{x_1}^{x_2} E_n \left( t, \frac{y_2 - y_1}{x_2 - x_1} (t - x_1) + y_1 \right) \\ &\quad \cdot \sqrt{1 + \left( \frac{y_2 - y_1}{x_2 - x_1} \right)^2} dt. \end{aligned} \quad (12)$$

Let  $(y_2 - y_1)/(x_2 - x_1) = k$ ,  $(y_1 x_2 - y_2 x_1)/(x_2 - x_1) = b$ , and when the  $P_1, P_2$  are both above the  $x$ -axis shown in Figure 6, formula (12) could be simplified as follows:

$$\begin{aligned} E_n(n-1) &= (v_n + v_\epsilon) \sqrt{k^2 + 1} \\ &\times \left\{ \int_{x_1}^{x_2} \frac{1}{\sqrt{t^2 + (kt + b)^2}} dt \right. \\ &\quad \left. + (1 - \alpha) \int_{x_1}^{x_2} \frac{kt + b}{t^2 + (kt + b)^2} dt \right\}. \end{aligned} \quad (13)$$

If the points  $P_1$  and  $P_2$  are below the  $x$ -axis shown in Figure 6, we could get formula (14)

$$\begin{aligned} E_n(n-1) &= (v_n + v_\epsilon) \sqrt{k^2 + 1} \\ &\times \left\{ \int_{x_1}^{x_2} \frac{1}{\sqrt{t^2 + (kt + b)^2}} dt \right. \\ &\quad \left. + (1 - \alpha) \int_{x_1}^{x_2} \frac{-kt - b}{t^2 + (kt + b)^2} dt \right\}. \end{aligned} \quad (14)$$

If the  $P_1$  and  $P_2$  are located across the  $x$ -axis as Figure 7 shows, assume the interaction point between  $P_1P_2$  and  $x$ -axis is  $P_3(x_3, 0)$ , and let  $x_1 < x_3 < x_2$ . It is needed to calculate the field intensity of  $P_1P_3$  and  $P_3P_2$  separately, as follows:

$$\begin{aligned} E_n(n-1) &= (v_n + v_\epsilon) \sqrt{k^2 + 1} \\ &\times \left\{ \int_{x_1}^{x_3} \frac{1}{\sqrt{t^2 + (kt + b)^2}} dt \right. \\ &\quad + (1 - \alpha) \int_{x_1}^{x_3} \frac{kt + b}{t^2 + (kt + b)^2} dt \\ &\quad + \int_{x_3}^{x_2} \frac{1}{\sqrt{t^2 + (kt + b)^2}} dt \\ &\quad \left. + (1 - \alpha) \int_{x_3}^{x_2} \frac{-kt - b}{t^2 + (kt + b)^2} dt \right\}. \end{aligned} \quad (15)$$

The prerequisite of the formula (15) is  $x_1 < x_2$ . And if  $x_1 = x_2$ , let  $x_1 = x_2 = a$  and assume that

$$x = a, \quad y = t. \quad (16)$$

The field intensity of segment  $P_1P_2$  is shown as follows:

$$\begin{aligned} \int_{P_1P_2} E_n(x, y) ds &= (v_n + v_\epsilon) \sqrt{k^2 + 1} \\ &\times \left[ \int_{y_1}^{y_2} \frac{1}{\sqrt{a^2 + t^2}} dt + \int_{y_1}^{y_2} \frac{|t|}{a^2 + t^2} dt \right]. \end{aligned} \quad (17)$$

And if the  $P_1$  and  $P_2$  are located across the  $x$ -axis as Figure 7 shows, it is needed to calculate the field intensity of  $P_1P_3$  and  $P_3P_2$  separately, as follows:

$$\begin{aligned} \int_{P_1P_2} E_n(x, y) ds &= (v_n + v_\epsilon) \sqrt{k^2 + 1} \\ &\times \left\{ \int_{y_1}^{y_3} \frac{1}{\sqrt{a^2 + t^2}} dt + \int_{y_1}^{y_3} \frac{t}{a^2 + t^2} dt \right. \\ &\quad \left. + \int_{y_3}^{y_2} \frac{1}{\sqrt{a^2 + t^2}} dt + \int_{y_3}^{y_2} \frac{-t}{a^2 + t^2} dt \right\}. \end{aligned} \quad (18)$$

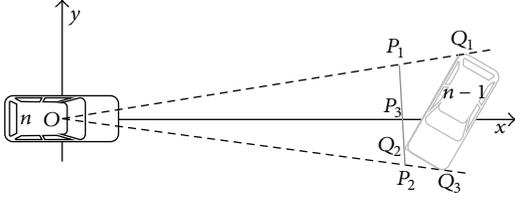


FIGURE 7: Object projection in driver's visual range schematic II.

Thus, it can extend this method to calculate any form of the object. Because the projection of the object could finally be attributed to a segment in attention field. And if the segment endpoints is known, it is easy to get the field intensity  $E_n(P_1P_2)$  of the segment than getting that of the object.

**3.2.2. The Attention Field Range.** Limited by the human eye vision scope, the driver's attention field is finite. According to general cognitive, with the vehicle speed increasing, the driver's attention distance expands in the direction of traveling. But limited by the driver's visual capacity, the attention distance exits upper limit. Combined with previous, take the max value of upper limit to be 420 meter, and the min upper limit to be 50 meter. So assume the attention range in the traveling direction is shown as formula (19). The unit of  $v_n$  is km/h.  $R$  is the maximum visual attention distance with the speed  $v_n$ ; its unit is meter:

$$R = \begin{cases} 60 & v_n < 40, \\ 6v_n - 180 & v_n \in [40, 100], \\ 420 & v_n > 100. \end{cases} \quad (19)$$

The driver's visual range would be gradually narrow with the vehicle speed increasing. Assume the attention field range is shown as in Figure 8, the points  $L_{l1}$ ,  $L_{l2}$ ,  $L_{r1}$ , and  $L_{r2}$  are the four boundary points of the attention field effect range. Let the angle between  $L_{l1}O$  and  $L_{r1}O$  be  $\eta$ . The relationship between  $\eta$  and vehicle speed is shown in Table 1. According to the data in Table 1, establish the corresponding math relationship between the speed and the visual angle  $\eta$  with the regression least squares method as follows:

$$\eta = 0.0064v_n^2 - 1.8244v_n + 159.55. \quad (20)$$

In the transverse direction, the driver usually only focuses on the objects in the driveway and the adjacent two-lane and neglects the objects outside this range. Assume the length  $L_{l2}L_{r2}$  in Figure 8 is  $d_y$ ; let it equal to triple lane width as formula (21) shows. The  $d_{\text{lane}}$  is the lane width. And the coordinates of the four boundary points in the attention field effect range could be obtained. And the attention field effect scope is identified:

$$d_y = 3d_{\text{lane}}. \quad (21)$$

**3.2.3. Effect of Object Motion Direction.** There would be different impacts of the objects in the same spatial position but

TABLE 1: Correspondence between speed and visual angle.

Vehicle speed (km/h)	0	40	70	100
Visual angle (degree °)	160	95	65	40

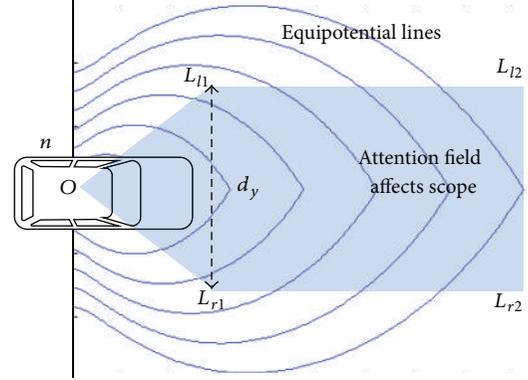


FIGURE 8: Attention field affects scope schematic.

with different relative moving direction to the target vehicle driver. In model, this embodies the field intensity changing.

In physical, the gradient describes the steepest changing direction of the field intensity. Because of the discontinuities of the attention field on the  $x$ -axis, using exhaustive method to calculate the direction with the field intensity greatest changing point is needed.

Take the specific point  $(x, y)$  as the center, and the specific distance  $r$  as the radius, calculate the field strength on the circumference.  $r$  is positively relative with the relative velocity. Compared with the field strength at point  $(x, y)$ , the connection between the  $(x, y)$  and the biggest field intensity changing point is the direction of gradient in attention field; let it be  $Ga(x, y)$ .

Meanwhile, we could get the angle between the  $Ga(x, y)$  and direction of the relative movement object speed; let it be angle  $\varphi$ . If the relative speed coincides with  $Ga(x, y)$ , it indicates that the object would have a relatively big impact on driver. And with the angle  $\varphi$  increasing, the psychological pressure of the driver from the object get weakened.

**3.2.4. The Effect of Psychological Drive.** Driving task brings the driver psychological pressure. In the free state, the pressure comes from the driver's desired speed. The driver would accelerate by adjusting the speed and finally stabilize nearby the expectation. In the car-following state, the drive force comes from the large driving distance to the leading car. In order to ensure traffic efficiency, the driver accelerates to maintain the proper headway or the distance to the leading vehicle. The psychological drive force and the psychological pressure from the attention field alternation affect the driver's driving behavior.

### 3.3. Numerical Simulation

**3.3.1. The Free State.** When the vehicle travels in the free state, the surrounding vehicles sparse; the driver's mainly

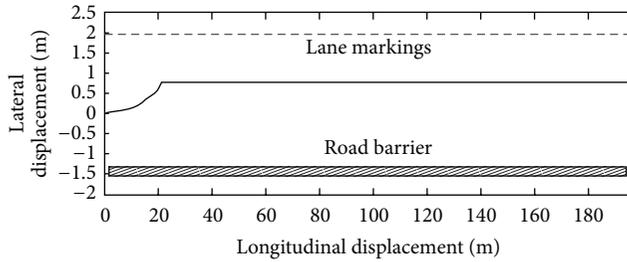


FIGURE 9: Schematic of vehicle running track.

restrictions are road conditions and constraints from the traffic facilities. The lane marking and the road barrier are the most common driving traffic facilities. The road surface conditions determine the driver's desired speed, and lane markings and road barrier constrain the direction of the vehicle trajectory. Analyze the driving behavior in a free state on road as Figure 9 shows.

Take the impacts of driver from lane markings and road barrier as a simple vector superposition. Assume the vehicle running in the road as Figure 9 shows without any other vehicles and obstacles. Let the lane width be 3.5 meters; the distance between the vehicle and the road barrier in the perpendicular to the traveling direction is 1.5 meters at initial time. The initial velocity and the expected velocity are both 20 m/s.

Apply the model this paper proposed to simulate the vehicle running processing. The displacement curve is shown in Figure 9. In initial simulate time, the vehicle is close to the road barrier. For keeping the safety, the vehicle decelerated and shifted away from the road barrier in initial time stage. When the offset reached a certain extent, limited by the lane line in Figure 9, the vehicle stopped moving in the transverse direction. At this time, because the distance between the road barrier and the vehicle in vertical direction reached a safety extent, the driver maintained the vehicle running state.

In the two-lane road, according to the observed situation in traffic environment, the vehicle is not usually located in the middle of the lane, but has a certain extent offset in the vertical direction of road. This phenomenon is due to psychological stress caused by the road edge. In order to avoid this kind of pressure, the driver would leave away the road edge to reduce the attention to it if the road condition is allowed, thereby forming the above-described phenomenon.

**3.3.2. The Waiting Pedestrian Interference.** It is a normal phenomenon that the pedestrian crosses the road with no traffic signal controlling. As shown in Figure 10, the pedestrian waiting at roadside A for the gap of the traffic flow to cross the road reaches another roadside B. In this process, in addition to the crossing pedestrian affecting the vehicle in traffic flow, the ones in the waiting area also impact the driver. Then achieve the numerical simulation of this phenomenon.

The waiting pedestrians mostly would be cluster. For the cluster pedestrians, their projection in the driver's attention field is shown as the line  $Q_1Q_2$  in Figure 11. Based on the

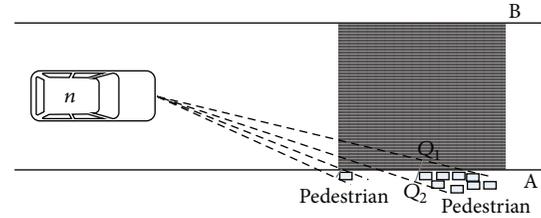


FIGURE 10: Pedestrian crossing schematic.

attention field introduction, the effect of the cluster pedestrian relative to the area and shape of cluster pedestrian. Assume the individual pedestrian occupies  $0.5 \times 0.5 \text{ m}^2$  space. Let the number of the pedestrians be 30, array in  $10 \times 3$  matrix; the first row of the crowd is at the roadside A as Figure 10 shows. Let the running vehicles queue number be 50. The first vehicle of the queue has a 50 m distance to the pedestrians. The initial and expected velocities are both 20 m/s. The space headway in the queue is 40 m.

The vehicle would slow down when it gets close to the pedestrian. Because the vehicle of the queue is in the car-following state, the decelerate behavior would spread back in the queue, causing the whole velocity of the traffic flow to reduce. And when the vehicle passes the pedestrians, it would be affected by both the leading vehicle and the pedestrians, if the leading just pass the pedestrian effect range and begin to accelerate, the following vehicle driver would not be able to speed up with the leading as usual, but continue to be affected by the pedestrian. Until leaving the pedestrian effect range, the vehicle accelerates to the expected speed with the psychological driving force. So in Figure 11(a), the velocity of the vehicles queue would first decrease and then recover; the density of the queue declines continuously after the initial slight increase. And maintains stable after leaving the effected range.

It could be seen from Figure 11(b) that, after the vehicles queue passing the pedestrian effect range, the queue would be stretched. This is because the first vehicle passes the pedestrians effect range with reducing the speed, the slowdown state spread back to the queue. And when the first vehicle just passed the effect range, the expected speed drives the driver to accelerate to the initial state. But this state would not spread immediately because the subsequent vehicles are still affected by the pedestrian. The space headway of the neighboring vehicles increases. And limited by the expected velocity, even the subsequent vehicles have passed the effect range, they would not accelerate continuously but maintain an expected speed. So if the smooth running traffic flow passes through the pedestrian effect area, the spacing between the neighboring vehicles and the density of the traffic flow decrease, causing the number of passing vehicles in unit time to reduce, and the traffic capacity of the road reduces.

## 4. Conclusion

This paper presents a new method of modeling the driving behavior. Used the field concepts in physical to establish

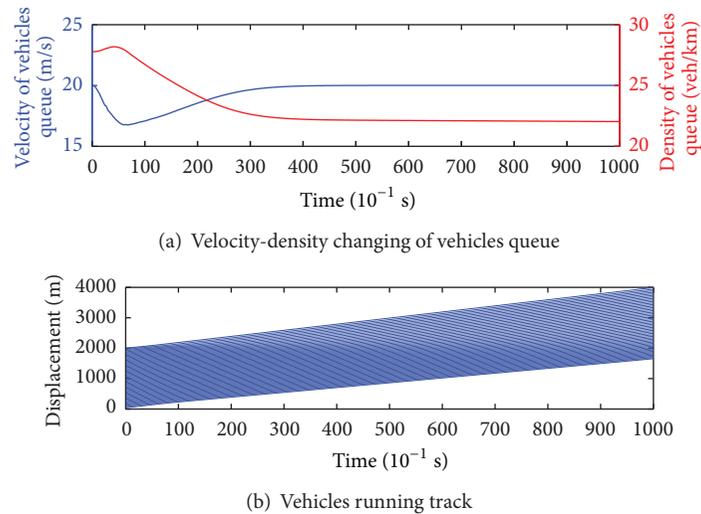


FIGURE 11: Displacement of the vehicles in queue.

a driver's attention field to describe the distribution of the attention. The result shows that the model can simulate the operation of driver process, and in line with the individuals vehicle movement and the traffic flow running.

The model extends the scope of the driving behavior model, getting rid of the limited by car-following behavior. And the model could output the vehicle in the lateral movement. Meanwhile, because of the complex of the driving behavior, the model is not capable of describing all driving behaviors. The model structure needs to be further restructured and strengthened. And more experimental data is needed to specifically calibrate the parameters to make the model applied in practice.

## Conflict of Interests

The author declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

This work is supported by the National Science Foundation of China (nos. 51278220, 51278454, and 51108208).

## References

- [1] M. Brackstone and M. McDonald, "Car-following: a historical review," *Transportation Research F*, vol. 2, no. 4, pp. 181–196, 1999.
- [2] A. Kesting and M. Treiber, "Calibrating car-following models by using trajectory data methodological study," *Transportation Research Record*, no. 2088, pp. 148–156, 2008.
- [3] Y. Hollander and R. Liu, "The principles of calibrating traffic microsimulation models," *Transportation*, vol. 35, no. 3, pp. 347–362, 2008.
- [4] D. C. Gazis, R. Herman, and R. W. Rothery, "Nonlinear follow the leader models of traffic flow," *Operations Research*, vol. 9, no. 4, pp. 913–933, 1961.
- [5] S. Jin, D.-H. Wang, C. Xu, and Z.-Y. Huang, "Staggered car-following induced by lateral separation effects in traffic flow," *Physics Letters A*, vol. 376, no. 3, pp. 153–157, 2012.
- [6] B. Gunay, "Car following theory with lateral discomfort," *Transportation Research B*, vol. 41, no. 7, pp. 722–735, 2007.
- [7] S. Jin, D. Wang, P. Tao, and P. Li, "Non-lane-based full velocity difference car following model," *Physica A*, vol. 389, no. 21, pp. 4654–4662, 2010.
- [8] H. B. Zhu and S. Q. Dai, "Analysis of car-following model considering driver's physical delay in sensing headway," *Physica A*, vol. 387, no. 13, pp. 3290–3298, 2008.
- [9] S. Ossen and S. P. Hoogendoorn, "Heterogeneity in car-following behavior: theory and empirics," *Transportation Research C*, vol. 19, no. 2, pp. 182–195, 2011.
- [10] H. Rakha and M. Arafeh, "Calibrating steady-state traffic stream and car-following models using loop detector data," *Transportation Science*, vol. 44, no. 2, pp. 151–168, 2010.
- [11] R. Hoogendoorn, S. Hoogendoorn, K. Brookhuis, and W. Daamen, "Psychological elements in car-following models: mental workload in case of incidents in the other driving lane," *Procedia Engineering*, vol. 3, no. 1, pp. 87–99, 2010.
- [12] P. Wagner, "Analyzing fluctuations in car-following," *Transportation Research B*, vol. 46, no. 10, pp. 1384–1392, 2012.
- [13] F. E. Gunawan, "Two-vehicle dynamics of the car-following models on realistic driving condition," *Journal of Transportation Systems Engineering and Information Technology*, vol. 12, no. 2, pp. 67–75, 2012.
- [14] W. Wang, K. Bengler, and G. Wets, "Discrete dynamics in transportation system," *Discrete Dynamics in Nature and Society*, vol. 2013, Article ID 234970, 2 pages, 2013.
- [15] H. Guo, W. Wang, W. Guo, and F. Zhao, "Modeling lane-keeping behavior of bicyclists using survival analysis approach," *Discrete Dynamics in Nature and Society*, vol. 2013, Article ID 197518, 6 pages, 2013.

## Research Article

# A Study on High-Speed Rail Pricing Strategy in the Context of Modes Competition

Enjian Yao,<sup>1,2</sup> Qirong Yang,<sup>1,2</sup> Yongsheng Zhang,<sup>1,2</sup> and Xun Sun<sup>1,2</sup>

<sup>1</sup> State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup> School of Traffic and Transportation, Beijing Jiaotong University, Haidian District, Beijing 100044, China

Correspondence should be addressed to Qirong Yang; 12120935@bjtu.edu.cn

Received 30 September 2013; Accepted 5 December 2013

Academic Editor: Huimin Niu

Copyright © 2013 Enjian Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

High-speed rail (HSR) has developed rapidly in China over the recent years, for the less pollution, faster speed, comfort, and safety. However, there is still an issue on how to improve the seat occupancy rates for some HSR lines. This research analyzes the pricing strategy for HSR in Wuhan-Guangzhou corridor based on the competition among different transport modes with the aim of improving occupancy rates. It starts with the theoretical analysis of relationship between market share and ticket fare, and then disaggregate choice models with nested structure based on stated preference (SP) data are established to obtain the market share of HSR under specific ticket fare. Finally, a pricing strategy is proposed to improve the occupancy rates for Wuhan-Guangzhou HSR. The results confirm that a pricing strategy with floating fare should be accepted to improve the profit of HSR; to be specific, the ticket fare should be set in lower level on weekdays and higher level on holidays.

## 1. Introduction

High-speed rail (HSR) is currently regarded as one of the most significant technological breakthroughs in passenger transport developed in the second half of the 20th century [1]. Due to the advantages of rapidness, comfort, convenience, safety, and reliability [2], China has witnessed rapid development of HSR over the past years. However, its performance in operation is still restricted by the pricing strategy under the intense competition among various transport modes. The traditional fixed pricing strategy gives up the induced passenger flow generated by fares change [3] and the high pricing of HSR leads to low occupancy rates and resources waste. For example, statistics shows that the average occupancy rates of Wuhan-Guangzhou HSR can be as low as 20% except for the spring festival which is not satisfactory as expected. The lower the occupancy rate is, the less the profit is. Therefore reasonable pricing strategy should be researched to solve the pricing problems for HSR.

In order to solve the problem of pricing strategy, a number of recent researchers have been devoted to studying

reasonable methodology for passenger transport pricing. Li and Tayur [4] and Labbé et al. [5] applied the Bilevel programming to the optimal pricing. Zeng et al. [6] put forward a new thought of combining the value of travel time and Bilevel programming to maximize the benefit of the railway agencies and the passengers' utility. Hsu et al. [7] and Adler et al. [8], based on game theory, analyzed the competition between two modes of transport and get optimal pricing in order to maximize the profits of operator. Zhou et al. [9] studied the pricing model for parallel rail lines under the situation of diversified property rights through considering the main influencing factors of the rail network pricing, including cost and supply.

Though extensive researches have been undertaken to search for optimal passenger transport pricing, few researchers have been devoted to studying the relationship between ticket fare and market share of transport mode. However, the demand (market share) for certain mode changes along with passenger pricing policy [10]; that is, certain mode price variation will affect the market share while market share variation will affect ticket fare in case of maximizing

TABLE I: Mode split in the sample for different income levels.

	Sample size				Total	The ratio of income level (%)
	HSR	Conventional rail	Air	Road		
Low-income group (lower than 30,000 CNY)	529	207	275	11	<b>1022</b>	33.20
Middle-income group (between 30,000 CNY and 100,000 CNY)	710	204	507	10	<b>1431</b>	46.49
High-income group (higher than 100,000 CNY)	279	56	287	3	<b>625</b>	20.31

operators' profits. Therefore, ticket fare and market share cannot be separated from each other.

This paper analyzes a pricing strategy for high speed rail (HSR) based on the quantitative relationship between ticket fare and market share. The rest of this paper is organized as follows. In the second section, nested choice models for different income levels are established. The model parameters are estimated and individual preferences are analyzed in the third section. The fourth section studies the quantitative relationship between market share and ticket fare, and then a pricing strategy aimed to improve the occupancy rates of HSR is proposed. The conclusions are given in the fifth section.

## 2. Nested Choice Model

Disaggregate choice analyses, based on SP (stated preference), RP (revealed preference), or mixed data, are usually advocated by researchers as a proper methodology to assess and compare the preferences of travelers in the context of model competition [11]. To analyze the market share of HSR, disaggregate choice models based on the SP information provided by the survey are estimated for this study.

**2.1. The Data.** A questionnaire survey in Wuhan-Guangzhou corridor was conducted to obtain stated preference (SP) data for the model estimation. Questionnaires were distributed to public transport users in railway station, airport, and so on. The SP data was obtained by presenting 9 profiles, in which the attributes of HSR such as travel time and travel cost were varied. And the attributes of current alternatives were left unchanged. In each profile, respondents were asked to make a choice from the given alternatives: HSR, conventional rail, air, and road transport. Besides, personal information was also asked for in the SP survey such as age, profession, trip purpose, and income.

A total of 3078 valid observations were obtained from the questionnaire survey. Considering that the sensitivities to multiple attributes are different under various income levels, the obtained data can be divided into three datasets according to annual income: low-income group, middle-income group, and high-income group. And three models with different datasets are established, respectively. The distribution of income levels and the mode split in the sample are described in Table 1.

From the available information in the sample, HSR has an absolute advantage in attracting passengers and the new alternative will capture passenger flow from existing modes.

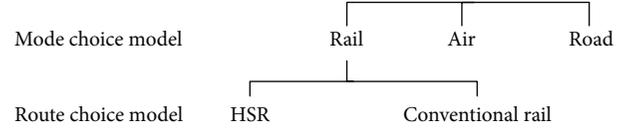


FIGURE 1: Structure of nested mode/route choice model.

**2.2. Model Formulation.** Multinomial logit (MNL) model is the traditional and popular tool used among logit models. However, MNL model exhibits the independence from irrelevant alternative (IIA) property so that it fails to account for the existence of similarities among choice alternatives [12]. The nested logit (NL) model overcomes the problem by grouping alternatives into nests, and interdependence between the pairs of alternatives is allowed in the same layer to satisfy the IIA property [13–15]. A nested choice model is considered and the model structure is shown in Figure 1. The travel modes considered in this study are rail, air, and road. The rail mode is divided into two alternatives: HSR and conventional rail.

Disaggregate choice model has theoretical basis on the assumption of utility maximization. The probability that individual  $n$  chooses alternative  $i$  is given by

$$\begin{aligned}
 P_{in} &= P(U_{in} > U_{jn}, i \neq j) \\
 &= P(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn}, i \neq j),
 \end{aligned} \tag{1}$$

where  $U_{in}$  is the utility of alternative  $i$  for individual  $n$ ;  $V_{in}$  is the deterministic term in the utility function of alternative  $i$  for individual  $n$ ;  $\varepsilon_{in}$  presents the random term.

When the random terms obey the distribution of Gumble, the probability that individual  $n$  chooses an alternative is shown from (2) to (4). The probability of mode  $m$  being chosen can be calculated by (2). The conditional probability of choosing route  $r$  given that mode  $m$  is chosen can be described by (3). Furthermore, (4) gives the probability that a route  $r$  is chosen. Consider the following:

$$P_n(m) = \frac{e^{\lambda(V_{m^*} + V_{m^*}^*)}}{\sum_{m'=1}^{M_n} e^{\lambda(V_{m'^*} + V_{m'^*}^*)}}, \tag{2}$$

$$P_n(r | m) = \frac{e^{(V_{r|m})_n}}{\sum_{r'=1}^{R_{m^*}} e^{(V_{r'|m^*})_n}}, \tag{3}$$

$$P_n(rm) = P_n(r | m) P_n(m), \tag{4}$$

TABLE 2: Dummy variables used in the utility function.

	HSR	Conventional rail	Air	Road
<b>Age</b>				
Over forty	1	0	0	0
Other	0	0	0	0
<b>Profession</b>				
Civil servants or managers	1	0	1	0
Workers	0	1	0	0
Other	0	0	0	0
<b>Trip purpose</b>				
Business purpose	1	0	1	0
Other	0	0	0	0

where  $\lambda$  is the scale parameter;  $M_n$  are the set of modes that exist in mode  $m$ ;  $R_{mn}$  are the set of alternatives that exist in route  $r$ ;  $V_{(r|m)_n}$  is the fixed term in the utility function that varies with the combination of  $m$  and  $rm$ ;  $V_{mn}$  is the fixed term in the utility function that has nothing to do with  $r$  and only varies with  $m$ ;  $V_{mn}^*$  is the utility composited based on the fixed term of  $rm$ . And the composed utility can be described in

$$V_{mn}^* = \ln \sum_{r=1}^{R_{mn}} \exp(V_{(r|m)_n}). \quad (5)$$

For each of the utility functions in the three models, we take account for some attributes such as travel time, travel cost, profession, age, and trip purpose. As for the service attributes of transport modes (i.e., travel time and travel cost), the same parameter is applied to travel time of HSR and air and the same parameter is used to travel cost of HSR and air too. Given the airport location, a terminal time is taken into account in the utility of air transport. To take advantage of other attributes, it is essential to express the attributes as concrete numbers when used in the utility function. Through analyzing individual preferences of the sample, it is assumed that passengers over forty and those traveling for business purpose have a general preference for HSR and passengers with a variety of professions show different preferences for certain mode. For example, civil servants and managers prefer to choose HSR or air for travelling. On the contrary, workers are more willing to choose conventional rail. The dummy variables used in the utility function are shown in Table 2.

### 3. Estimation Results

With various influence factors considered in the utility function, maximum likelihood estimations for different income groups are presented in Tables 3, 4, and 5. Some conclusions can be summarized by analyzing and comparing the results in different tables.

- (1) The facts that all absolute  $t$ -values are greater than 1.96 indicate that, for all coefficients, we can reject

the null hypothesis that the true value is zero at the 0.05 significance level. Meanwhile, the likelihood ratio indexes for all models are over 0.2 which can be regarded as satisfactory goodness of fit.

- (2) For all models, parameters of travel time and travel cost have a negative impact on the utility function. This is consistent with common sense that passengers try their best to reduce travel time and cost when traveling. Besides, parameters of travel time keep increasing from Table 3 to Table 5, which shows that passengers become more sensitive to the variation of travel time as the income level rises.
- (3) From the estimation results of all models, the parameters of profession, age, and trip purpose have a positive impact on the utility function. It means that HSR is very attractive for passengers traveling for business purpose as well as those in old age. Civil servants and managers show a general preference for HSR and air. On the contrary, workers prefer to choose conventional rail for travelling.
- (4) As for the value of travel time (VOTT) of HSR, represented by the single ratio between travel time and travel cost, it appears that the VOTT keeps increasing as the income level varies. The results indicate that high-income passengers are willing to pay more money in exchange for the decrease of travel time when they choose HSR for travelling.

## 4. High-Speed Rail Pricing

In this section, a pricing strategy with the aim of improving occupancy rates for Wuhan-Guangzhou HSR is discussed. Firstly, the quantitative relationship between ticket fare and market share of HSR can be obtained based on the dataset calculated through calibrated NL models, and then the pricing strategy is presented through considering the different passenger flow between weekdays and holidays.

*4.1. The Relationship between Market Share and Ticket Fare.* Based on the NL models with parameters calibrated, the HSR market share under specific fare can be calculated by (6). And it is composed of the market share of low-income group, middle-income group, and high-income group (i.e.,  $P_{\text{hsr}}^1$ ,  $P_{\text{hsr}}^2$ , and  $P_{\text{hsr}}^3$ ). And the market shares of different income groups can be obtained respectively from the following:

$$P_{\text{hsr}} = \frac{1}{i} \sum_{i=1}^3 \theta_i P_{\text{hsr}}^i, \quad (6)$$

$$P_{\text{hsr}}^1 = \frac{1}{n} \sum_{n=1}^K P_n^{\text{low}'}, \quad (7)$$

$$P_{\text{hsr}}^2 = \frac{1}{n} \sum_{n=1}^K P_n^{\text{mid}'}, \quad (8)$$

$$P_{\text{hsr}}^3 = \frac{1}{n} \sum_{n=1}^K P_n^{\text{high}'}, \quad (9)$$

TABLE 3: Estimations of low-income group ( $t$ -statistics are in parentheses).

	HSR	Conventional rail	Air	Road
Constant			2.0170 (5.326)	
Travel time (h)	-0.4972 (-6.105)	-0.3671 (-2.684)	-0.4972 (-6.105)	-0.3671 (-2.684)
Travel cost (CNY/100)	-0.8435 (-6.431)	-0.9741 (-2.055)	-0.8435 (-6.431)	-0.9741 (-2.055)
Profession	0.6348 (2.489)	0.6348 (2.489)	0.6348 (2.489)	
Age	0.4877 (3.664)			
Logsum		1.5026 (5.763)		
$\rho^2$		0.2846		
Sample size		1022		
VOTT (CNY/h)	58.94			

TABLE 4: Estimations of middle-income group ( $t$ -statistics are in parentheses).

	HSR	Conventional rail	Air	Road
Constant			2.3909 (6.881)	
Travel time (h)	-0.7663 (-8.914)	-0.4649 (-3.718)	-0.7663 (-8.914)	-0.4649 (-3.718)
Travel cost (CNY/100)	-1.0534 (-8.885)	-1.4410 (-3.208)	-1.0534 (-8.885)	-1.4410 (-3.208)
Profession	0.3680 (2.329)	0.3680 (2.329)	0.3680 (2.329)	
Trip purpose	0.3809 (2.583)		0.3809 (2.583)	
Logsum		1.3068 (8.266)		
$\rho^2$		0.3368		
Sample size		1431		
VOTT (CNY/h)	72.75			

TABLE 5: Estimations of high-income group ( $t$ -statistics are in parentheses).

	HSR	Conventional rail	Air	Road
Constant			3.3348 (4.809)	
Travel time (h)	-1.0467 (-6.205)	-0.4139 (-2.114)	-1.0467 (-6.205)	-0.4139 (-2.114)
Travel cost (CNY/100)	-1.2753 (-5.490)	-2.6309 (-3.746)	-1.2753 (-5.490)	-2.6309 (-3.746)
Trip purpose	0.7519 (3.963)			
Logsum		1.3077 (5.694)		
$\rho^2$		0.4255		
Sample size		625		
VOTT (CNY/h)	82.07			

where  $\theta_i$  is the weight measured by the distribution of income levels shown in Table 1;  $P_n^{\text{low}}$  is the probability of choosing HSR for a low-income passenger;  $P_n^{\text{mid}}$  is the probability of choosing HSR for a middle-income passenger;  $P_n^{\text{high}}$  is the probability of choosing HSR for a high-income passenger.

To analyze the relationship between market share and ticket fare, 41 ticket fares are picked from 390 CNY to 590 CNY in order. And the later fare is increased by 5 CNY than the former one. Then the market shares under specific ticket fares are calculated. Based on the regression analysis method, the quantitative relationship is shown in Figure 2. The adjusted  $R^2$  is up to 0.997, which indicates that logarithmic regression model is suitable to describe the relationship between the parameters. Given that the market share is a variable between 0 and 1, the ticket fare should range from 294 CNY to 673 CNY to make sure that the equation is effective. The relation equation is shown in (10) and

the value range of  $x$  is described in the parentheses. Consider the following:

$$y = -1.21 \ln(x) + 7.879, \quad (294 < x < 673), \quad (10)$$

where  $y$  is the HSR market share;  $x$  is the HSR ticket fare, CNY.

**4.2. Pricing Strategy.** Based on the quantitative relationship obtained, a pricing strategy to improve the occupancy rates between Wuhan and Guangzhou is researched. It is assumed that the rate of passenger transport demand from Wuhan to Guangzhou ranges from 20% to 35% in the corridor. With the aim of 100% occupancy rate, the pricing strategy for HSR is shown in Table 6.

From the estimation results, 100% occupancy rate can be implemented when the fare is set at 407 CNY in weekdays and 533.5 CNY (or 420 CNY) in holidays. In weekdays, the fare

TABLE 6: Estimations of HSR pricing.

Operation scheme	Weekdays		Holidays
	8 carriages, 43 trains	8 carriages, 43 trains	16 carriages, 43 trains
Assumption of travel demand (day)	8000	30000	30000
HSR pricing (CNY)	407	533.5	420

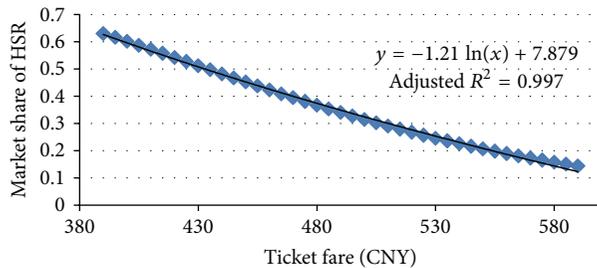


FIGURE 2: Regression analysis results.

is 50 CNY lower than the current one (465 CNY), so it will be beneficial to decrease the fare appropriately in exchange for the increase of occupancy rates. In holidays, the operation scheme in weekdays even cannot meet the transport demand. But when the carriages of each train are increased to 16, 420 CNY should be the satisfying fare for obtaining the 100% seat occupancy rate. Comparatively speaking, a strategy with floating pricing should be more positive to improve the earnings. Ticket fare should be set in lower level in order to attract more passengers in weekdays. As for the pricing strategy in holidays, both the rise of ticket fare and the adjustment in operation scheme could be effective.

## 5. Conclusions

This paper provides a pricing strategy for Wuhan-Guangzhou HSR based on the quantitative relationship between rail pricing and market share of HSR. Through considering the service attributes of transport mode and personal attributes, NL models using SP data are built to obtain the market share of HSR under specific fare. This method not only suits for assessing and comparing the individual preferences under the context of mode competition but also gives a pricing strategy to relieve the situation of low occupancy rates for some HSR lines. The results of nested choice model confirm that the sensitivities to multiple influencing factors are diverse as income level varies and passengers with high income pay more attention to the travel time of transport mode other than personal properties. Besides, the results of pricing strategy show that floating ticket fare will be more positive to improve the occupancy rates for HSR and to meet transport demand. The pricing strategy obtained could be beneficial to the full play of economic and social benefits under the rapid development of HSR in China.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of the paper.

## Acknowledgments

This work is supported by State Key Laboratory of Rail Traffic Control and Safety (no. RCS2011ZT012) and National 973 Program of China (no. 2012CB725403).

## References

- [1] J. Campos and G. de Rus, "Some stylized facts about high-speed rail: a review of HSR experiences around the world," *Transport Policy*, vol. 16, no. 1, pp. 19–28, 2009.
- [2] L. Wang, G. Wu, and Z. Zhang, "Literature review and prospects for pricing of high-speed railroad," in *Proceedings of the 3rd International Conference on Transportation Engineering (ICTE '11)*, pp. 1349–1353, Chengdu, China, July 2011.
- [3] X. Zhang, W. X. Luan, Q. D. Cai et al., "Research on dynamic pricing between high speed rail and air transport under the influence of induced passenger flow," *Information Technology Journal*, vol. 11, no. 4, pp. 431–435, 2012.
- [4] L. Li and S. Tayur, "Medium-term pricing and operations planning in intermodal transportation," *Transportation Science*, vol. 39, no. 1, pp. 73–86, 2005.
- [5] M. Labbé, P. Marcotte, and G. Savard, "A bilevel model of taxation and its application to optimal highway pricing," *Management Science*, vol. 44, no. 12, pp. 1608–1622, 1998.
- [6] J. Zeng, Y. J. Liu, and Y. Y. Lin, "Pricing model of train passenger transport based on the value of travel time and bi-level programming," in *Proceedings of the 2013 20th International Conference on Management Science & Engineering*, pp. 393–399, Harbin, China, 2013.
- [7] C. Hsu, Y. Lee, and C. Liao, "Competition between high-speed and conventional rail systems: a game theoretical approach," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3162–3170, 2010.
- [8] N. Adler, E. Pels, and C. Nash, "High-speed rail and air transport competition: game engineering as tool for cost-benefit analysis," *Transportation Research B*, vol. 44, no. 7, pp. 812–833, 2010.
- [9] S. N. Zhou, Q. S. Zhang, and X. W. Wu, "A new pricing model of China's parallel rail lines under the diversified property rights," *Journal of Industrial Engineering and Management*, vol. 6, no. 1, pp. 135–148, 2013.
- [10] M. Wardman, "Demand for rail travel and the effects of external factors," *Transportation Research E*, vol. 42, no. 3, pp. 129–148, 2006.
- [11] C. Roman, R. Espino, and J. C. Martin, "Analyzing competition between the high speed train and alternative modes. The case

- of the Madrid-Zaragoza-Barcelona corridor,” *Journal of Choice Modelling*, vol. 3, no. 1, pp. 84–108, 2010.
- [12] C. Wen, W. Wang, and C. Fu, “Latent class nested logit model for analyzing high-speed rail access mode choice,” *Transportation Research E*, vol. 48, no. 2, pp. 545–554, 2012.
- [13] E. Yao, T. Morikawa, S. Kurauchi, and T. Tokida, “A study on nested logit mode choice model for intercity high-speed rail system with combined RP/SP data,” in *Proceedings of the 3rd International Conference on Traffic and Transportation Studies (ICTTS '02)*, pp. 612–619, July 2002.
- [14] E. Yao and T. Morikawa, “A study of an integrated intercity travel demand model,” *Transportation Research A*, vol. 39, no. 4, pp. 367–381, 2005.
- [15] M. C. J. Bliemer, J. M. Rose, and D. A. Hensher, “Efficient stated choice experiments for estimating nested logit models,” *Transportation Research B*, vol. 43, no. 1, pp. 19–35, 2009.

## Research Article

# Modeling of a Small Transportation Company's Start-Up with Limited Data during Economic Recession

Xiaoping Fang,<sup>1</sup> Jake Ansell,<sup>2</sup> and Weiya Chen<sup>1</sup>

<sup>1</sup> School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

<sup>2</sup> Business School, University of Edinburgh, Edinburgh EH8 9JS, UK

Correspondence should be addressed to Weiya Chen; [wychen@csu.edu.cn](mailto:wychen@csu.edu.cn)

Received 9 October 2013; Revised 11 November 2013; Accepted 25 November 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Xiaoping Fang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a modeling method for analyzing a small transportation company's start-up and growth during a global economic crisis which had an impact on China which is designed to help the owners make better investment and operating decisions with limited data. Since there is limited data, simple regression model and binary regression model failed to generate satisfactory results, so an additive periodic time series model was built to forecast business orders and income. Since the transportation market is segmented by business type and transportation distance, a polynomial model and logistic curve model were constructed to forecast the growth trend of each segmented transportation market, and the seasonal influence function was fitted by seasonal ratio method. Although both of the models produced satisfactory results and showed very nearly the same of goodness-of-fit in the sample, the logistic model presented better forecasting performance out of the sample therefore closer to the reality. Additionally, by checking the development trajectory of the case company's business and the financial crisis in 2008, the modeling and analysis suggest that the sample company is affected by national macroeconomic factors such as GDP and import & export, and this effect comes with a time lag of one to two years.

## 1. Introduction

Transport infrastructure is critical to economic development of a country and can provide competitive advantage. Within China there is a diversity of transport companies which could be split between with large-scale transport enterprises (LTEs) and small- and medium-sized transportation enterprises (SMTEs). There are many distinct differences between the two types of enterprises. Generally, LTEs, more geographically spread, attract larger enterprises and resource management is not as critical. SMTEs especially in their start-up period need to be more agile in the use of their financial resources to ensure survival. This is particularly true during a period of economic change. This paper focuses on SMTEs since they are the main suppliers of road transportation services in China. They necessarily play the role of first-and/or-last kilometer carriers for door-to-door logistics services. Some 225 259 enterprises and 4 595 600 individually owned businesses were engaged in transportation, storage,

and postal services in 2008 [1]. The total value of the top 50 logistics enterprises, about 475.6 billion Yuan, is only 0.53% of the national added value of the logistics industry [2]. Compared with LTEs, SMTEs face much greater risk, especially during an economic downturn. The survival of a business entity depends heavily on its ability to anticipate and prepare for change rather than wait for it and then react to it [3].

Demand forecasting is so significant for SMTEs that it can help to improve equipment utilization and establish smarter operational and investment strategies. Realistically, business owners estimate the service demand from their past experiences which can be wrong or misleading. Almost all freight demand analysis usually accounts for a whole country, a region, or a corridor between cities by integration or by mode and is usually to do with public transportation planning [4–6]. Two previous research articles which examined SMTEs' demand forecasting tackled less than truckload (LTL) for both short-term and long-term forecasting. In [7],

TABLE 1: Business data by industry.

Industry	No. of customers		No. of orders		Money	
	Absolute value	Percentage %	Absolute value	Percentage %	Absolute value	Percentage %
Food	1	3.125	24	0.2221	91900	0.4495
Toy	1	3.125	154	1.4253	430492	2.1054
Chinaware	1	3.125	358	3.3133	949029	4.6414
Material	2	6.250	151	1.3975	363945	1.7799
Machine	3	9.375	4423	40.9348	4474611	21.8837
Craft	4	12.500	1001	9.2642	2867729	14.0250
Electronics	5	15.625	1999	18.5007	2541727	12.4307
Logistics	15	46.875	2695	24.9422	8727782	42.6845
Total	32	100	10805	100	20447215	100

it is believed that the combination of neural networks and traditional time series analysis is good for forecasting short-term logistic demand for an LTL carrier. A small-to-medium-sized road transport service which offered a limited collection and delivery service for small and large consignments in a number of geographical zones of equal area is reported in [8].

The objective of this study is to model the start-up and growth of a newly established truck transportation company during the economic recession whose main business is sea-port containers and bulk inland transportation. It ultimately aims to help SMTEs to look for ways of improving equipment utilization in the short to medium term by forecasting the orders and the trucks required, as well as the impact of financial crisis. The main feature of our work is its use of limited data to analyze and forecast a small transportation company's business as it starts up in a rapidly changing environment.

The context of the model has two distinct elements: enterprise start-up and the short time series. Obviously, most enterprises in their start-up period need to grow their business to an economic scale. If they fail to do so the long time survival of the enterprise may be under question. Hence, the model selected will have to accommodate such growth. The second element reflects the need for quick appraisal of likely demand, so that appropriate economic strategy can be developed by the enterprise. This differentiates start-up SMTEs from LTEs where there is possibly longer times series and higher correlation of performance with macroeconomic factors. The modeling proposed attempts to cope with these two elements.

## 2. Data

The data covers all the business orders of a small truck transportation company during a 40-month period (January 2008 to April 2011). Registered at the end of 2007, the small truck transportation company currently has a fleet of 30 trucks and 30 drivers. Some trucks are container trucks suitable for container transportation and others are bulk trucks designed for heavy long-distance bulk transportation services. This company mainly provides services for pier container transportation and inland bulk transportation. The container transportation service is mainly provided

for container customers who have exporting and importing businesses at three ports located in Shenzhen, namely, Yantian, Shekou, and Mawan. In order to make use of the empty containers at these ports, a small part of the bulk is transported for the container customers in container trucks at the same price as bulk transportation. Most of the rest is carried by bulk trucks for different customers.

All the business orders from this company are classified according to the categories of the customer industry (see Table 1). "Money" in this paper represents the gross income which equals the orders multiplying the unit price. The customers come from eight industries, and 15 logistics customers contribute 42.7% of the total money, three machine customers account for 40.9% of the orders and 21.9% of the gross income, and the remaining 14 customers from the food, toy, chinaware, material, craft, and electronics industries account for 34.123% of the orders and 35.432% of the total gross income.

Further analysis shows that there are two patterns as regards gaps in orders, regular and irregular. Regular orders are made either continuously on weekdays and broken at weekends and vacations (shown in Figure 1(a)) or discretely and regularly (shown in Figure 1(b)). Irregular orders are made randomly, as shown in Figures 2(a) and 2(b). The customers who make regular and irregular orders are listed in Table 2.

In order to find out whether the economic recession of 2008 had an effect on the company's business, further analysis of the orders and money per month was conducted for both regular and irregular businesses. It should be noted that the Chinese New Year usually falls in February when almost all businesses are at their lowest level of production. A three-year period is considered and the cycle starts from February of the first year to January of the last year. Therefore, "1" in Figure 3 represents February 2008 and "36" means January 2011.

The plots of the irregular orders and money per month shown in Figure 3 describe two factors. (1) The irregular business was affected by the economic recession which hit China between the third and fourth quarters of 2008. The downturn began in October 2008. (2) Orders declined rapidly from October 2008 and eventually disappeared.

Originating in the developed countries, the economic crisis soon spread to China, as shown by the decreasing number

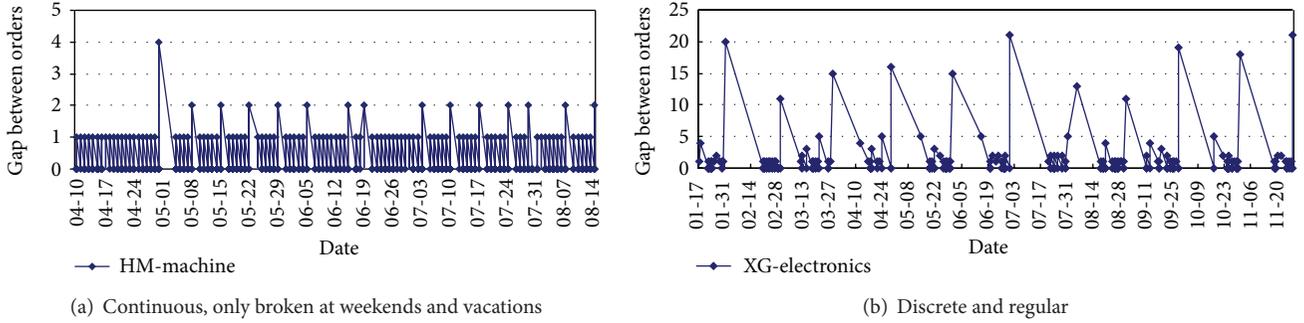


FIGURE 1: Gaps between regular orders in days.

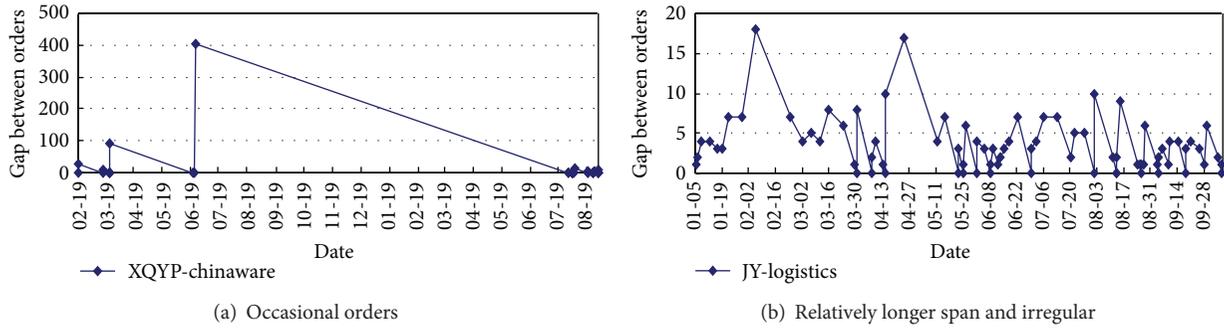


FIGURE 2: Gaps between irregular orders in days.

TABLE 2: Regular and irregular customers.

Regular: 12 customers, 9147 orders Name/industry		Irregular: 20 customers, 1658 orders Name/industry	
HYu/craft	YLM/toy	OWY/craft	QS/logistics
GL/craft	HM/machine	HS/craft	XQ/logistics
XG/electronics		HYa/electronics	YLX/logistics
GD/electronics		WLi/electronics	CH/logistics
WLin/electronics		ZH/food	GY/logistics
GH/logistics		LH/logistics	JFX/logistics
HF/logistics		LR/logistics	FY/logistics
KLD/logistics		QFT/logistics	XQ/chinaware
KZ/logistics		JY/logistics	SH/machine
HL/material		AD/material	HLD/machine

of orders. Those companies involved in overseas markets were affected first, in the third quarter of 2008. Pure OEM (original equipment manufacturing) and excessive reliance on overseas market were the two major factors which made these companies fail in the recession rapidly. Those companies relying on the home market were affected by a time lag, because of reduced business with the export-oriented companies and the declining purchasing power of workers laid off by those companies.

The regular orders and money per month in total are given in Figure 4, from which conclusions can be drawn as follows.

- (1) There is obvious periodicity by year both for orders and for income.
- (2) There is an increase in trend, but the slope is gradually decreasing.
- (3) Orders and money surprisingly increase after the arrival of the recession.

Some weak companies ceased trading almost immediately, but the robust ones survived. Businesses concentrated on those surviving companies.

Every coin has two sides. The recession had both negative and positive impacts on the company's business. Whereas the business trade was slowing down, the quality of the company's customers was better after the recession and a large number of irregular businesses disappeared. Graphs of the orders and income have similar shapes, so the regular group is the focus of the following work.

### 3. Modeling and Forecasting

Most modeling and forecasting approaches with limited data are about rapidly changing industries like motion pictures, telecommunications, or new products with a short history [9–13]. They suggest that combining ARIMA and diffusion models can improve one-year-ahead predictions, especially in the high technology market. The drawback, however,

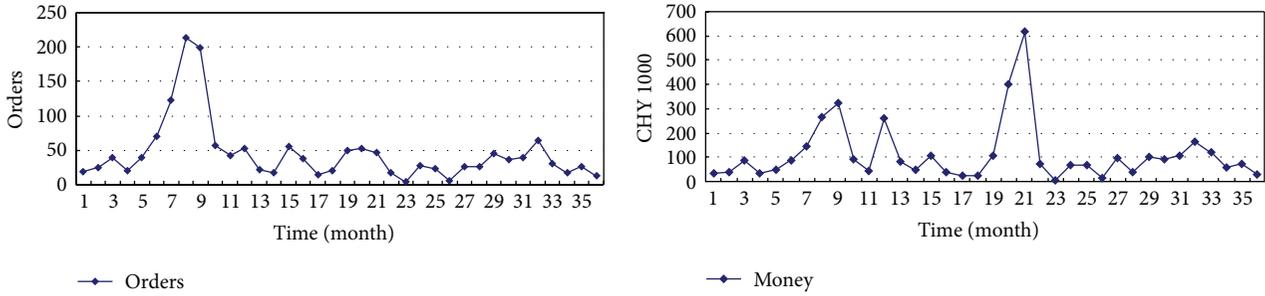


FIGURE 3: Irregular orders and money per month in total.

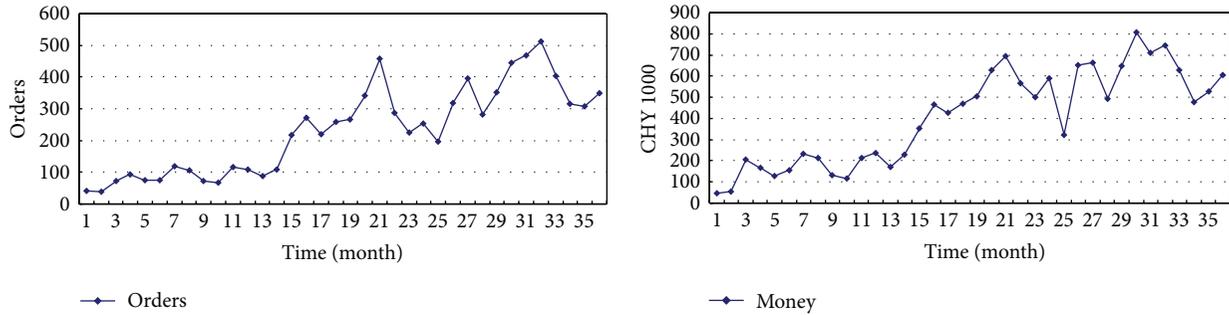


FIGURE 4: Regular orders and money per month in total.

is the need for enough historical data points to create a time series. The necessity of judgment in forecasting with insufficient data for statistical methodology was discussed in [14]. The key is to impose a structure with methods such as surveys of intentions or expectations, judgmental bootstrapping, structured analogies, and simulated interaction. A case study about fashion products, generally characterized by high demand uncertainty, high stock out costs, and a high risk of obsolescence, proposes that preorder data and judgments can be used to overcome the lack of data in demand forecasting. Advance order data can be obtained by allowing a selected group of customers to preorder at a discount from a preview catalogue. Judgments can be obtained from purchase managers or other company experts [15].

A significant improvement in nonlinear time series analysis (NTSA) has been seen since the 1990s [16]. Tong listed the following as the five most promising directions: the interface between NTSA and chaos, the nonparametric/semiparametric approach, nonlinear state space modeling, financial time series (in both discrete time and continuous time), and nonlinear modeling of panels (e.g., spatially distributed) in time series. A polynomial spline approximation was employed to estimate the functional coefficient regression models for nonlinear time series [17]. A time-varying coefficient time series model with a time trend function and serially correlated errors to characterize the nonlinearity, nonstationarity, and trending phenomena was discussed in [18]. A time series can be decomposed into trend, cycle, and seasonal and random components [19]. That is, moving averages, isolating seasonal factors and seasonal

adjustments are common methodology for treating seasonal adjustment of time series.

No single model or combination model has been proved to outperform the rest. In order to model and forecast the small truck transportation company's start-up and growth after the economic recession from the limited data described in Section 2, relative variable analysis and time series analysis were used to predict the orders and turnover. Meanwhile, correlation analysis, and polynomial and logistic curve models were also employed to find the trend by combining seasonality adjustment to predict order number and turnover.

Simple linear regression and binary linear regression are two basic versions of the generalized linear model (GLM) proposed by [20] as a way of unifying various other statistical models, including linear regression, logistic regression, and Poisson regression.

Linear regression was employed to search for the relation between orders and import & export in months or orders and import & export and GDP in quarters for GDP only available in quarters. The general form of a linear regression model is

$$y = a + b_i x_i + e \quad i = 1, 2, \dots, n, \quad (1)$$

where  $y$  is the dependent variable, denoting the orders.  $x_i$  ( $i = 1, 2, \dots, n$ ) is the predictor variable with a subscript  $i$ , and  $n$  is the number of variables.  $e$  is noise.

In statistics, signal processing, econometrics, and mathematical finance, a time series is a sequence of data points, measured typically at successive times at uniform time intervals. Time series forecasting is the use of a model to predict future values based on previously observed values. Methods

for time series analyses may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and, recently, wavelet analysis; the latter include autocorrelation and cross-correlation analysis.

Our time series analysis began by regrouping the data, splitting the distance into five sections and business into two kinds. The first model was the polynomial, as in the following:

$$O_{ds} = \mu + \beta_1 t_y + \beta_2 t_y^2 + \beta_3 t_y^3 + s_i. \quad (2)$$

Then for an alternative of the logistic curve regression

$$O_{ds} = \alpha_{ds} \cdot \left( \frac{e^{\beta_{ds} t_m}}{1 + e^{\beta_{ds} t_m}} \right). \quad (3)$$

Therefore, the turnover forecast is as the following for both trend forecast models:

$$\text{money} = \sum_{d=1}^5 \sum_{s=1}^2 V_{ds} \cdot O_{ds}, \quad (4)$$

where  $V_{ds}$  is average price or value of distant  $d$  and type  $s$ ,  $O_{ds}$  the orders of distant  $d$  and type  $s$ ,  $S_i$  the noise,  $t_y$  is the time by year, and 1 = the starting year, 2008 in this case.  $\alpha_{ds}$  and  $\beta_{ds}$  are coefficients of the logistic model in subgroups of a distance  $d$  ( $= 1, 2, 3, 4, 5$ ) and a type  $s$  ( $= 1, 2$ ).  $t_m$  ( $= 1, 2, \dots, 36$ ) represents the time by month.

Any time series is made up of systematic components, such as a trend, cycle, and seasonal and random elements, which are by definition unpredictable. Therefore, a set of time series data can be decomposed into trend, cycle, and seasonal and random components. The four elements can be combined in either an additive or a multiplicative model [19]:

$$X_{t_m} = T + C + S + R \quad (5)$$

$$X_{t_m} = T \times C \times S \times R, \quad (6)$$

where  $X_{t_m}$  represents the dependent variable of orders,  $T$  the trend component,  $C$  the seasonal factor, and  $R$  the random element.

Moving averages are often used to isolate the trend. We propose, however, to extract the trend directly from the original data with both the polynomial and logistic curve models and then to decompose the combination of the cycle and seasonality by the additive model in (5). Considering the close relationship between transportation and economics, we can ignore the cyclical component by taking the seasonality as a cycle of 12 months. The cycle was previously defined as running from February to January of the next year, as the lowest demand usually falls in February.

We take the additive model, (5), and rewrite it as

$$X_{t_m} - T = S + R. \quad (7)$$

Let  $SR$  ( $= S + R$ ) be the combination of the seasonal and random components and then divide  $SR$  by the trend  $T$  to get the seasonal ratio  $S_c$  (including the random). Hence, the seasonal factor can be obtained by averaging the three ratio

TABLE 3: Segmentation of the distance.

$d$	1	2	3	4	5
distance	<90 km	90–179 km	180–350 km	351–800 km	>801 km

components (February 2008 to January 2009, February 2009 to January 2010, and February 2010 to January 2011) for each month  $\bar{S}_c$ ,  $c = 1, 2, \dots, 12$ . For example, for February, the average seasonal ratio  $\bar{S}_c$  is as follows:

$$\bar{S}_1 = \frac{1}{3} \left( \left. \frac{X_{t_m} - T}{T} \right|_{t_m=1} + \left. \frac{X_{t_m} - T}{T} \right|_{t_m=13} + \left. \frac{X_{t_m} - T}{T} \right|_{t_m=25} \right), \quad (8)$$

where  $t_m$  ( $= 1, 2, \dots, 36$ ) represents months from the beginning in the series. Now the forecasting value of  $X_{t_m}$  can be obtained by

$$\hat{X}_{t_m} = T + S + R = T \left( 1 + \bar{S}_c \right) \quad (9)$$

$$t_m = 1, 2, \dots, 36; \quad c = 1, 2, \dots, 12.$$

Then the total orders in (10) and (4) should be changed to

$$\text{orders} = \sum_{d=1}^5 \sum_{s=1}^2 O_{ds}^{tr} \left( 1 + \bar{S}_c \right) \quad (10)$$

$$\text{money} = \sum_{d=1}^5 \sum_{s=1}^2 V_{ds} \cdot O_{ds}^{tr} \left( 1 + \bar{S}_c \right). \quad (11)$$

$O_{ds}^{tr}$  is the trends forecast by the polynomial or logistic model.

## 4. Application

First, simple linear regression analysis is taken to model the relationship between the national export & import and the orders. The result is given in Figure 5.

The predicted regular orders per month =  $-148.296 + 0.330 * \text{national import \& export}$ .

Then the binary linear regression of the orders and imports & exports and GDP per quarter is presented in Figure 6.

The predicted regular orders per quarter =  $-560.432 + 1.076 * \text{GDP} + 0.158 * \text{import \& export}$ .

As the relative analysis cannot give satisfactory results even intuitively, we turn to the time series analysis. The whole regular business is separated by distance and truck type when the time series analysis is chosen. Distance is separated into five categories according to the rates and transport time, and container and bulk are the two types of services. Variable  $d$  represents the different segmentation of the distance shown in Table 3. Type variable  $s = 1$  represents container and  $s = 2$  the bulk.

The result shows that almost all the container transport orders are included in distance groups 1, 2, and 3, with just one out of 7755 regular container orders from February 2008 to January 2011 in container group 4. Most bulk transport orders are included in distance groups 1, 3, and 5. As the data of bulk

TABLE 4: The seasonal ratios in the polynomial model.

Group	Cycle											
	1	2	3	4	5	6	7	8	9	10	11	12
11	-0.4104	-0.3695	-0.1292	-0.0093	-0.1897	-0.1207	0.2280	0.3509	0.5158	0.0453	0.0209	-0.1597
21	-0.4775	-0.3617	-0.2347	0.1407	0.2389	0.3525	0.3019	0.2410	0.0508	-0.1913	0.0472	0.2199
31	-0.4325	-0.1774	0.5607	-0.1737	-0.1317	0.3161	0.2004	0.3484	-0.1109	-0.5404	-0.2736	0.3707
12	0.5909	-0.8788	0.1113	-0.0745	0.4005	0.3488	-0.3943	-0.1522	-0.0579	-0.1450	0.2265	0.2134
22	-0.6800	0.6150	0.3167	-0.1666	-0.4749	-0.0283	-0.4198	-0.3530	-0.1552	-0.0745	-0.4022	-0.2205

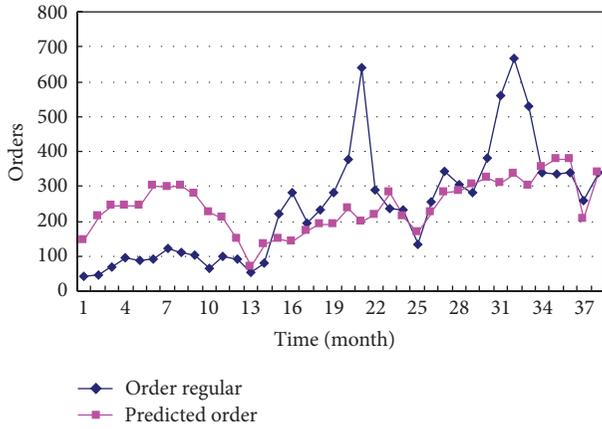


FIGURE 5: Simple linear regression of the regular orders and national import &amp; export.

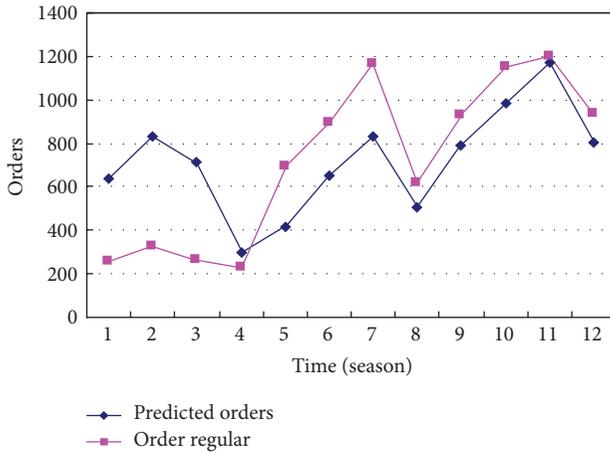


FIGURE 6: Binary linear regression: regular orders and import &amp; export and GDP.

groups 2 and 3 are insufficient for modeling, we merge bulk groups 1, 2, and 3 into a new bulk group 1 and bulk groups 4 and 5 into a new bulk group 2.

Thus,  $O_{41} = O_{51} = O_{32} = O_{42} = O_{52} = 0$ , and only the remaining five subgroups need to be estimated. The results of the polynomial trend models and seasonal ratios in Table 4

are listed in the following calculations and shown in Figures 7 and 8:

$$O_{11} = (-55.7778 + 91.58333t_y) \cdot (1 + \bar{S}_c^{11})$$

$$O_{21} = (10 + 27.45833t_y) \cdot (1 + \bar{R}_c^{21})$$

$$O_{31} = (4.9444 + 9.08333t_y) \cdot (1 + \bar{R}_c^{31})$$

$$O_{12} = (-1.98967 + 2.961088t_y^2 - 0.38809t_y^3) \cdot (1 + \bar{R}_c^{12}) \quad (12)$$

$$O_{22} = (-43.6667 + 58.0833t_y - 13.5833t_y^2) \cdot (1 + \bar{R}_c^{22}).$$

Forecast orders of all groups are presented in Figures 7 and 8.

Putting (12) into (11) we can calculate the money (turnover). The results show that (1) the container and the bulk have linear and quadratic trends, respectively and (2) the results are acceptable until March 2011, but the linear trend just increases and the quadratic trend abruptly decreases after April 2011, which is not in line with the actual situation.

Therefore, an alternative logistic model is needed. Equation (3) is the assumed trend function of orders. Let (10) and (11) be the total orders and money prediction formulation.

The order-prediction models are given in (13), with the second bulk group showing a polynomial trend:

$$O_{11} = 10.638 \times \frac{e^{0.95+0.104i}}{1 + 0.0519e^{0.95+0.104i}} \cdot (1 + \bar{R}_c^{11})$$

$$O_{21} = 1.51 \times \frac{e^{0.95+0.104i}}{1 + 0.0169e^{0.95+0.104i}} \cdot (1 + \bar{R}_c^{21})$$

$$O_{31} = 3.96 \times \frac{e^{0.95+0.104i}}{1 + 0.145e^{0.95+0.104i}} \cdot (1 + \bar{R}_c^{31}) \quad (13)$$

$$O_{12} = 12.823 \times \frac{e^{-1.605+0.1643i}}{1 + 0.164e^{0.95+0.104i}} \cdot (1 + \bar{R}_c^{12})$$

$$O_{22} = (-43.6667 + 58.0833t_y - 13.5833t_y^2) \cdot (1 + \bar{R}_c^{22}).$$

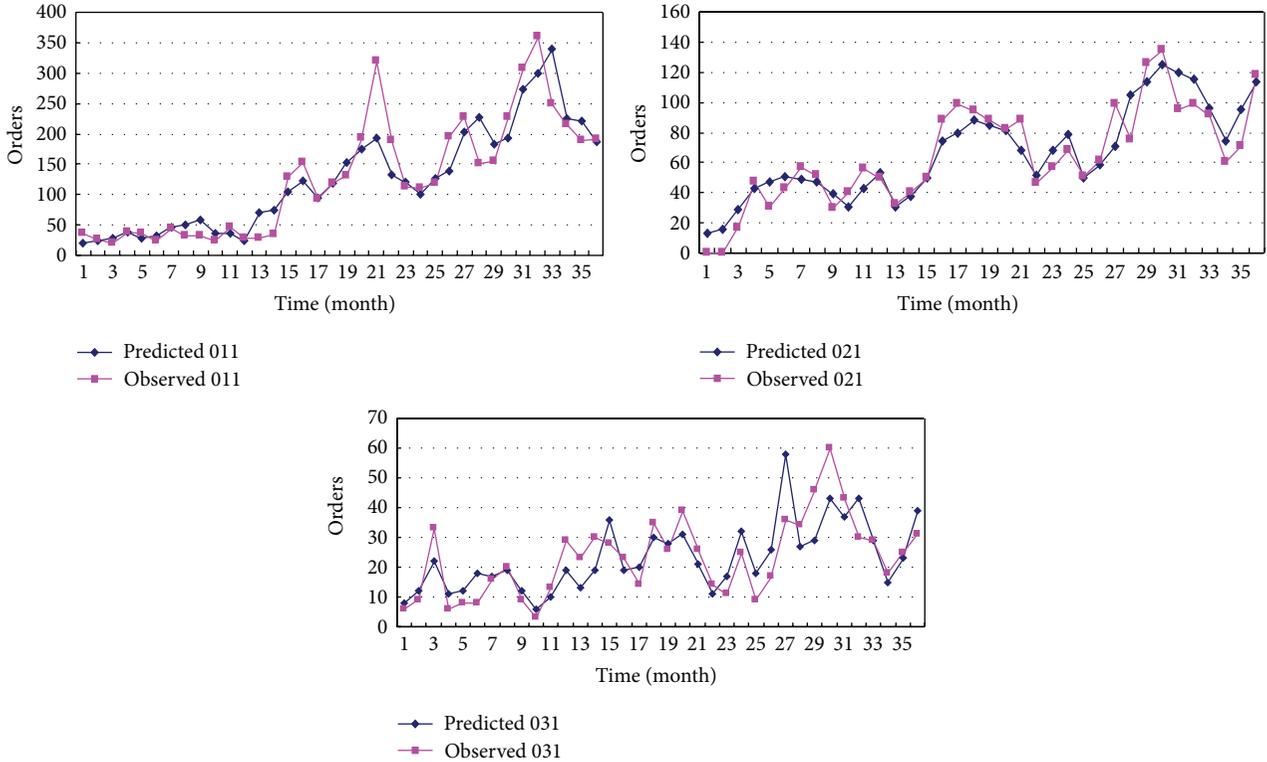


FIGURE 7: The orders of container transportation predicted by the polynomial model.

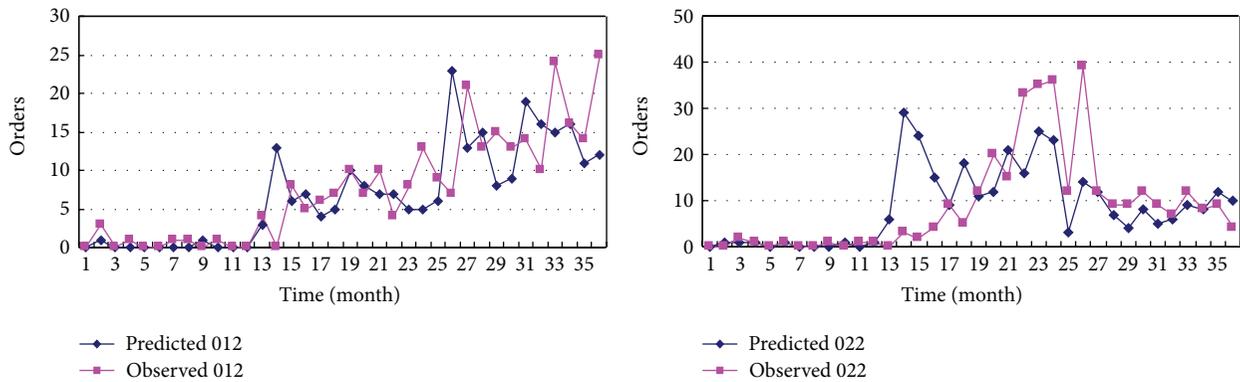


FIGURE 8: The orders of bulk transportation predicted by the polynomial model.

Results of calculation of seasonal ratios are presented in Table 5.

The results of the logistic model are shown in Figures 9 and 10. Only the group  $O_{22}$  could not be fitted by the logistic model.

The results of the polynomial and logistic models are compared in Figure 11. The predictions given by the two models are not very different at present but the logistic trend will be more significant in the future.

In conclusion, the initial modeling of the orders cannot immediately account for the demand in terms of exports & imports or GDP. The polynomial trend does not give a sound prediction in the sample, and the logistic modeling of

the orders seems appropriate. The forecasted total money is presented in Figure 12.

Model selection depends primarily on two aspects. One is the forecast in the sample. The second is the prediction out of the sample. The forecasts of the polynomial model decrease abruptly out of the sample and the logistic has a better performance in this aspect. Therefore the logistic model outperforms the polynomial model out of the sample.

Now let us turn to the analysis in the sample. There are some specific criteria, such as the coefficient of determination  $R^2$  and  $F$ -test, which measure the goodness-of-fit to discriminate which model is better than the others with regard to linear regression models. This is not the case with nonlinear

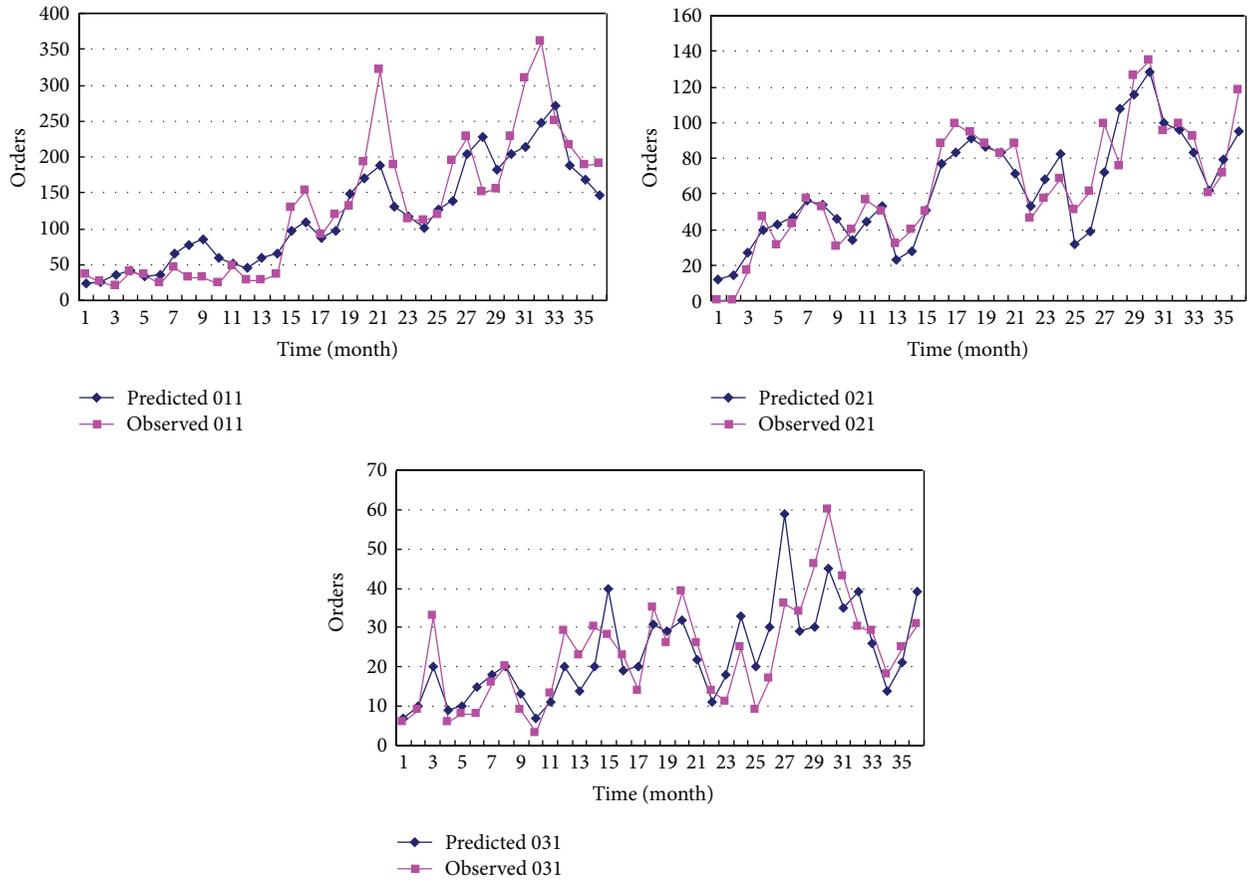


FIGURE 9: The container orders predicted by the logistic model.

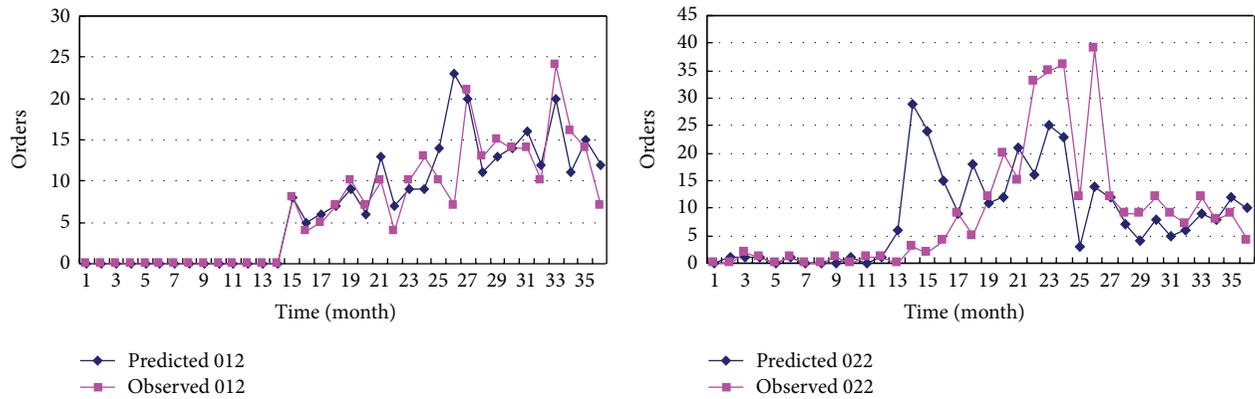


FIGURE 10: The bulk orders predicted by the logistic model.

TABLE 5: Calculation of seasonal ratios by the logistic model.

Group	Cycle											
	1	2	3	4	5	6	7	8	9	10	11	12
11	-0.1069	-0.0242	0.4384	0.6197	0.2914	0.4368	0.4109	0.6343	0.7970	0.2457	0.1089	-0.0297
21	-0.5951	-0.5055	-0.0923	0.3730	0.4713	0.6163	0.2290	0.1802	0.0183	-0.2400	-0.0248	0.1769
31	-0.1282	0.2625	1.5265	0.2267	0.2936	0.9351	0.4475	0.5893	0.0670	-0.4422	-0.1267	0.6150
12	0.3373	1.1651	0.6528	-0.0700	0.1107	0.0369	0.2534	-0.1151	0.4275	-0.1993	0.0686	-0.1443
22	-0.6800	0.6150	0.3167	-0.1666	-0.4749	-0.0283	-0.4198	-0.3530	-0.1552	-0.0745	-0.4022	-0.2205

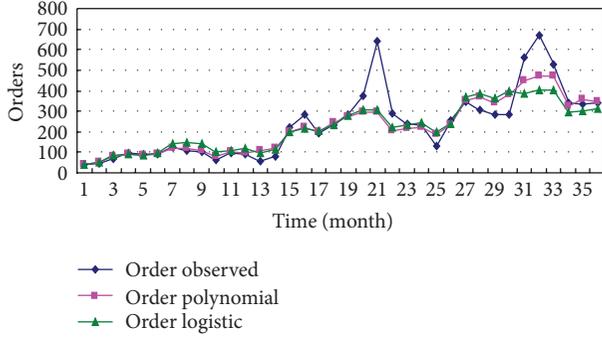


FIGURE 11: Comparison of the results of the two forecasting models: orders.

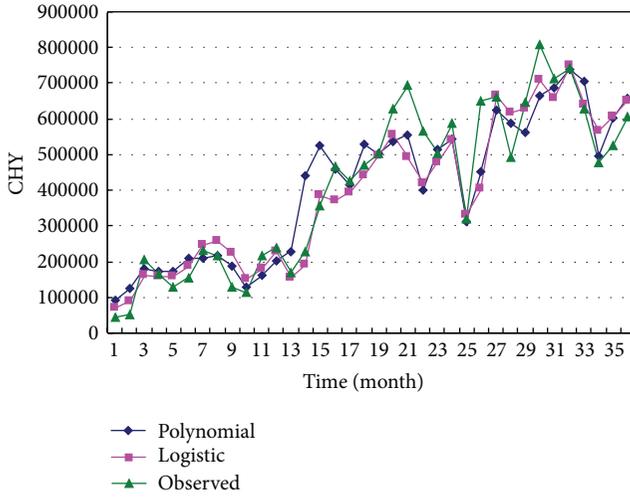


FIGURE 12: Comparison of the results of the two forecasting models: income.

regression models (NLRMs), however. According to [20], we cannot use the  $t$ -test (to test the significance of an individual coefficient) or the  $F$ -test (to test the overall significance of the estimated regression) because we cannot obtain an unbiased estimate of the error variance  $\sigma^2$  from the estimated residuals. Furthermore, the residuals (the difference between the actual  $y$  values and the estimated  $y$  values from the NLRM) do not necessarily sum to zero, ESS and RSS do not necessarily add up to the TSS, and therefore  $R^2 = \text{ESS}/\text{TSS}$  may not be a meaningful descriptive statistic for such models. An alternative of pseudo  $R$  square to  $\bar{R}^2$  is proposed by [21] as follows:

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (14)$$

where  $n$  is the observation number,  $y = \text{regress}$ , and  $\hat{u}_i = y_i - \hat{y}_i$ , where  $\hat{y}_i$  are the estimated  $y$  values from the (fitted) NLRM. Results are shown in Table 6.

From Table 6, the values of  $\bar{R}^2$  of orders prediction by the two models are nearly the same. When it comes to the income prediction, the value of  $\bar{R}^2$  of the logistic model has

TABLE 6: Measure of goodness-of-fit of polynomial and logistic trend models.

Model	Difference	Polynomial trend seasonal adjustment	Logistic trend seasonal adjustment
$\bar{R}^2$ /orders prediction	0.0066	0.7826	0.7760
$\bar{R}^2$ /money prediction	-0.0453	0.8569	0.9022

a slight advantage over the polynomial model. As previously described, the logistic model is better than the polynomial one in predicting performance out of the sample. The comparison in the sample shows, however, that performances of the two models are very similar. On balance, however, the logistic model is better than the polynomial.

### 5. Conclusions and Discussion

This paper introduces a small truck transportation company's start-up whose main business is pier containers and whose minority business is long-distance bulk transportation. We have attempted to elucidate how our analysis led to useful information for SMTE owners. Modeling the forecasting of order numbers of both kinds of businesses and the total turnover contributes to establishing the company's strategy for investment and operations.

Initial modeling of orders by simple regression and binary regression models cannot immediately account for demand in terms of exports or GDP. The results show that the company's business is certainly affected by the national macroeconomic factors such as GDP and import & export, and this effect comes with a time lag of one to two years since transportation service demand is derived demand. The seasonal fluctuation of orders, however, is much more dramatic than that of the national GDP and total import & export. It may be the main reason why the regression models did not perform well in forecasting. Competitive environment factors such as customers and competitors affect performance directly for SMTEs.

There are insufficient data about customers and competitors to support the simple regression model. Therefore, the time series analysis is the inevitable choice. We segmented the data by business and transportation distance to get a set of time series data. The polynomial and the logistic trend models combined with additional seasonal components were employed to fit the data of different segments. Both fitted the sample data very well, but the polynomial trend does not give sound forecasting for the fast decreasing trend out of the sample. From the perspective of goodness-of-fit, the polynomial trend model is still acceptable.

We have to say that forecasting in future is obviously at variance with reality. On the one hand, small transportation companies are not necessarily bound to continue to grow into large ones because of indifferent marketing and the lack of economies of scale. On the other hand, most small transportation companies can survive for a relatively long period thanks to their flexible operation mode. Usually a polynomial trend model does not work well for unlimited increasing or

decreasing trends while higher-order coefficients are positive or negative, respectively. One of the reasons why the logistic modeling of the orders seems appropriate is that it has given the upper limit of business. Acceptable goodness-of-fit and forecasting performance can only be provided by the logistic model, even though the performance of the two models in the sample is nearly the same to solve the question in this paper.

Unexpectedly, the analysis presented the process of the financial crisis's effect on SMTEs. First, irregular business disappears immediately the crisis emerges; at the same time, the regular business increases rather than decreases and then slows down with a time lag of one to two years, which is established by the logistic model. Within the effective distance of the truck transportation, the total transportation demand is relatively stable and the market is competitive or contestable. The growth rate of most small and medium transportation enterprises will become slower after their quick growth period and have an upper limit scale of business.

Our approach can cope with the different impetus of SMTEs' growth and a period of financial stress. One would expect that in less periods of financial stress then growth might be more rapid for SMTEs but not that they necessarily will continue to grow rapidly. With further data one could explore how later SMTEs might develop in more benign economic contexts. We believe, however, that the proposed model would still achieve a reasonable fit.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant no. 61203162. It was also supported in part by the Science Progress and Innovation Program of Hunan DOT under Grant no. 201244.

## References

- [1] National Bureau of Statistics of China, "The second economic census data bulletin (no.1)," 2009, <http://www.stats.gov.cn/tjsj/pcsj/jjpc/2jp/indexch.htm>.
- [2] National Development and Reform Commission, National Bureau of Statistics of China and China Federation of Logistics & Purchasing, "2008 National Logistics operation Situation Bulletin," 2009, [http://www.sdpc.gov.cn/jjyx/xdwl/t20090306\\_264999.htm](http://www.sdpc.gov.cn/jjyx/xdwl/t20090306_264999.htm).
- [3] X. Chen, X. Wang, and D. D. Wu, "Credit risk measurement and early warning of SMEs: an empirical study of listed SMEs in China," *Decision Support Systems*, vol. 49, no. 3, pp. 301–310, 2010.
- [4] J. Ansell and J. Banasik, "Forecasting white goods in a recession," *IMA Journal of Management Mathematics*, vol. 6, no. 3, pp. 315–331, 1995.
- [5] T. J. Fite, G. D. Taylor, J. S. Usher, J. R. English, and J. N. Roberts, "Forecasting freight demand using economic indices," *International Journal of Physical Distribution & Logistics Management*, vol. 32, no. 4, pp. 299–308, 2002.
- [6] S. Shen, T. Fowkes, T. Whiteing, and D. Johnson, "Econometric modelling and forecasting of freight transport demand in Great Britain," in *Proceedings of European Transport Conference*, <http://abstracts.aetransport.org/>.
- [7] L. Zhou, B. Heimann, and U. Clausen, "Forecasting a logistic service demand based on neural network," in *Proceedings of the International Conference on Service Systems and Service Management (ICSSSM '06)*, vol. 1, pp. 530–534, October 2006.
- [8] J. Harris, *Fuzzy Logic Applications in Engineering Science, Intelligent Systems, Control and Automation: Science and Engineering*, Springer, Dordrecht, The Netherlands, 2006.
- [9] B. G. S. Hardie, P. S. Fader, and M. Wisniewski, "An empirical comparison of new product trial forecasting models," *Journal of Forecasting*, vol. 17, no. 3-4, pp. 209–229, 1998.
- [10] P. McBurney, S. Parsons, and J. Green, "Forecasting market demand for new telecommunications services: an introduction," *Telematics and Informatics*, vol. 19, no. 3, pp. 225–249, 2002.
- [11] A. Mukherjee and V. Kadiyali, "Forecasting in rapidly changing environments: an application to the U.S. motion picture industry," Cornell University Johnson School Research Paper Series 10-07.
- [12] G. Madden, R. Cooper, and R. Fildes, "Theoretically-motivated long-term forecasting with limited data," Working Paper 240, The MIT Center for Digital Business, 2008.
- [13] C. Christodoulos, C. Michalakelis, and D. Varoutas, "Forecasting with limited data: combining ARIMA and diffusion models," *Technological Forecasting & Social Change*, vol. 77, no. 4, pp. 558–565, 2010.
- [14] K. C. Green and J. S. Armstrong, "Demand forecasting: evidence-based methods," Working Paper, University of Pennsylvania Department of Marketing, 2005.
- [15] J. Mostard, R. Teunter, and R. De Koster, "Forecasting demand for single-period products: a case study in the apparel industry," *European Journal of Operational Research*, vol. 211, no. 1, pp. 139–147, 2011.
- [16] H. Tong, "Nonlinear time series analysis since 1990: some personal reflections," *Acta Mathematicae Applicatae Sinica*, vol. 18, no. 2, pp. 177–184, 2002.
- [17] J. Z. Huang and H. Shen, "Functional coefficient regression models for non-linear time series: a polynomial spline approach," *Scandinavian Journal of Statistics*, vol. 31, no. 4, pp. 515–534, 2004.
- [18] Z. Cai, "Trending time-varying coefficient time series models with serially correlated errors," *Journal of Econometrics*, vol. 136, no. 1, pp. 163–188, 2007.
- [19] M. Barrow, *Statistics for Economics, Accounting and Business Studies*, Pearson Education, London, UK, 5th edition, 2009.
- [20] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society, Series A*, vol. 135, no. 3, pp. 370–384, 1972.
- [21] D. N. Gujarati, *Basic Econometrics*, McGraw-Hill, New York, NY, USA, 4th edition, 2003.

## Research Article

# Multinomial Logit Model of Pedestrian Crossing Behaviors at Signalized Intersections

Zhu-Ping Zhou,<sup>1</sup> Ying-Shun Liu,<sup>1</sup> Wei Wang,<sup>2</sup> and Yong Zhang<sup>2</sup>

<sup>1</sup> Department of Traffic Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

<sup>2</sup> School of Transportation, Southeast University, Nanjing 210096, China

Correspondence should be addressed to Ying-Shun Liu; [yingshun@mail.njust.edu.cn](mailto:yingshun@mail.njust.edu.cn)

Received 29 August 2013; Revised 12 November 2013; Accepted 25 November 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Zhu-Ping Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pedestrian crashes, making up a large proportion of road casualties, are more likely to occur at signalized intersections in China. This paper aims to study the different pedestrian behaviors of regular users, late starters, sneakers, and partial sneakers. Behavior information was observed manually in the field study. After that, the survey team distributed a questionnaire to the same participant who has been observed, to acquire detailed demographic and socioeconomic characteristics as well as attitude and preference indicators. Totally, 1878 pedestrians were surveyed at 16 signalized intersections in Nanjing. First, correlation analysis is performed to analyze each factor's effect. Then, five latent variables including safety, conformity, comfort, flexibility, and fastness are obtained by structure equation modeling (SEM). Moreover, based on the results of SEM, a multinomial logit model with latent variables is developed to describe how the factors influence pedestrians' behavior. Finally, some conclusions are drawn from the model: (1) for the choice of being late starters, arrival time, the presence of oncoming cars, and crosswalk length are the most important factors; (2) gender has the most significant effect on the pedestrians to be sneakers; and (3) age is the most important factor when pedestrians choose to be partial sneakers.

## 1. Introduction

Due to high population density, rapid urbanization, and lack of adherence to traffic regulations by both drivers and pedestrians, traffic accidents involving pedestrians have become a major safety problem all over the world, particularly in developing countries. For example, in 2009, 16683 pedestrians were killed in traffic crashes in China, representing 24.62% of all traffic fatalities [1]. However, 4092 pedestrians were killed in traffic crashes, accounting for only 12.10% of the fatalities sustained in police-reported motor vehicle crashes in the US in 2009 [2].

Crashes involving pedestrians are most likely to occur when pedestrians are crossing roads, especially crossing at signalized intersections. In China, more than 50% of pedestrian crashes occurred at signalized intersections [1, 3]. However, illegal pedestrian behavior is common and widespread in China. Yang et al. claimed that, in developing cities like Xi'an, if a pedestrian is waiting at a signalized intersection, in most cases they are waiting for an acceptable gap in

traffic and not for the green signal [4]. According to our study on the behaviors of 6628 pedestrians at 102 signalized crosswalks, the average compliance rate is only 62.8%. These statistics underlie the importance of understanding factors that contribute to pedestrian errors or willful violations of traffic regulations in China [5].

## 2. Previous Work

There have been considerable studies that have examined the factors which influence pedestrians' road crossing behaviors, such as traffic environment conditions, road user variables, and social factors.

*2.1. Traffic Environmental Conditions.* To identify the impact of urban planning on pedestrian crossings, a large study is carried out in Canada comparing pedestrian behaviors in Ontario with that in Quebec [6]. Using an attitude survey and a video survey, Keegan and O'Mahony studied the impact

of the pedestrian waiting countdown timer on pedestrian behavior [7]. They found that the timer units could induce a reduction in the number of individuals who crossed on red. In another more complete study, Chu et al. discussed the impact of street environment, including traffic conditions, roadway characteristics, and signal-control characteristics on the crossing behavior [8]. More recently, Tiwari et al. concluded that, as the signal waiting time increases, pedestrians are more likely to violate the traffic signal [9].

*2.2. Road User Variables.* Several studies have examined gender and age differences in pedestrian behaviors [10–12]. Male pedestrians tend to violate traffic rules more frequently than females. Generally, young adults and adolescent pedestrians are more likely to commit violations than older pedestrians (e.g., Moyano Díaz and Holland and Hill), and older road users express more appreciation for signalized intersections than younger pedestrians (e.g., Bernhoft and Carstensen) [12–14]. Other factors such as marital status, education, income, and personality variables such as attitude towards risk, sensation seeking and aggression were found to be related to pedestrian behavior [15–17].

*2.3. Social Factors.* The impact of others' behaviors on the individual has been investigated and has been found to be complicated. Santor et al. found that peer conformity was one of the strongest predictors of risky behaviors in adolescents [18]. Rosenbloom found that the presence of other pedestrians waiting at the crosswalk upon a pedestrian's arrival decreased the likelihood of crossing on red light [19]. Moreover, Zhou et al. presented a survey on 426 pedestrians that investigated the effect of conformity tendency on Chinese pedestrians. It was found that people who exhibited a greater tendency towards social conformity also had stronger crossing intentions in following other pedestrians than low conformity people [20].

Although considerable research has been done in recent years to identify the factors influencing pedestrian crossing behavior, limited studies on the effect of pedestrian's preference and attitude are reported before. Considering the studies described above, offering valuable information about pedestrian crossing behaviors at signalized intersections, this study aims to assess the effect of pedestrian-related factors and traffic factors on pedestrian behaviors.

### 3. Data

*3.1. The Affecting Factors.* In this study, the affecting factors are divided into two categories: pedestrian-related factors and traffic factors. Pedestrian-related factors refer to the features relative to the individual pedestrians, including demographic and socioeconomic characteristics (gender, age, career, education, and income), family characteristics (marital status, with or without children), trip characteristics (trip purpose, owning driver license or not), and attitude variables (latent variables). Traffic factors include the attributes of facilities (crosswalk length, signal length) and traffic conditions (arrival time, accompanied or alone, number of pedestrians

TABLE 1: Descriptive statistics for the samples.

Individual properties	Number	%
Gender		
Female	966	51.44
Male	912	48.56
Age		
<18	246	13.10
18–29	642	34.19
30–44	446	23.74
45–59	376	20.02
60+	168	8.95

waiting at the curb, and number of pedestrians crossing the street).

*3.2. Survey Design.* According to the signal use, pedestrians can be classified into four types: pedestrians who cross the street during the green signal (regular users), pedestrians who begin to cross when the signal is green but do not finish on green (late starters), pedestrians who cross during the red signal (sneakers), and pedestrians who cross part of the crosswalk during the red signal and then continue crossing during the green signal (partial sneakers) [21]. Essentially, this study mainly focuses on the factors relating to these four types of crossing behaviors, which are the alternatives of the logit model.

Behavior information was observed manually in field study. After obtaining the behavior data, the survey team distributed a questionnaire to the same participant that has been observed, to acquire detailed demographic and socioeconomic characteristics as well as subjective preference indicators. Also, the serial number method (giving the same number to the behavior recording and questionnaire) was applied to guarantee all the data to describe the same participant.

*3.3. Data Collection.* In the survey, a total of 3952 pedestrians on 32 crosswalks in Nanjing were observed, among whom, only 1970 (49.85%) agreed to do the questionnaire. The data required to build a logit model should include behavior information and other contributing factors. Totally, 1878 questionnaires remained after we eliminated the incomplete ones. Gender and age distributions of the complete samples are shown in Table 1.

Statistical analysis is applied to these 1878 effective samples, and the results of the crossing behaviors are shown in Table 2. Only 54.53% of the participants are legal crossing pedestrians. The ones that obey the rules spatially, but not temporally, account for 87.42% of the total number (few are partial jaywalkers and jaywalkers). For this condition, we do not conduct a deeper analysis.

### 4. Methodology

The objectives of this research are to examine the effect of various factors and to find out the most significant predictors

TABLE 2: Statistical results of the crossing behaviors.

Pedestrian counts	On-crosswalk		Partial jaywalkers		Jaywalkers		Total	
	Number	%	Number	%	Number	%	Number	%
Regular users	1024	54.53	4	0.21	8	0.42	1036	55.17
Late starters	206	10.97	34	0.85	18	0.96	240	12.78
Sneakers	270	14.38	38	2.02	70	3.74	378	20.13
Partial sneakers	142	7.54	16	1.81	48	2.56	224	11.92
Total	1642	87.42	92	4.89	144	7.68	1878	100.00

Note: “%” is a ratio of pedestrian counts of a certain type and the total sample (1878).

which influence the pedestrian’s crossing behaviors. The overall methodology to analyze how the factors affect pedestrian behavior involves three main steps.

*Step 1: Correlation Analysis to Identify the Factors.* Correlation analysis is needed to assess the actual effect of each factor relating to crossing behaviors, which is performed to identify the suitable variables for the discrete choice model. To analyze the correlation between the variables and behaviors, two different methods are applied separately for categorical variables and continuous variables. For the categorical variables, the crosstabs procedure is used, where single factor Pearson’s chi-square test is applied to determine whether a correlation exists between the factor and the crossing behaviors, while for the continuous variables, one-way variance analysis (ANOVA) is used for the same purpose as mentioned above. In the variance analyses,  $F$ -test is applied.

*Step 2: Structure Equation Modeling (SEM) to Link Pedestrians’ Preferences to Demographic Characteristics.* This step employs SEM, simultaneously estimating the relationships between the demographic data and the attitudinal factors and the relationship between the attitudinal statements and the attitudinal factors. SEM enables testing of a set of linear models to identify the structural attitudes of crossing behaviors and quantification of the causal relationships between pedestrians’ demographic profile and their attitudes [22]. Both manifest and latent variables are used in the SEM. There are two main groups of manifest variables: (1) the attitudinal indicator variables, which are the ratings that characterize pedestrians’ attitudes toward various crossings or travel statements and (2) socioeconomic and demographic variables, such as gender and income.

*Step 3: Multinomial Logit (MNL) Model.* The discrete choice model is the main model used in the study, which is based on the basic random utility theory assumption [23]. MNL model is developed to conduct the discrete choice analysis for three reasons. First, the study focuses on four types of pedestrian behaviors, and, thus, a multinomial model has to be used. Second, the model can handle a wide variety of variables (e.g., numerical, categorical). Third, because the MNL model is widely used, the empirical results can be easily interpreted and easily understood by decision makers who do not have a strong statistical background.

Traditionally, discrete choice models have considered only objective attributes from the alternatives and socio-economic characteristics of the individuals as explanatory variables. During the last decade, to capture the impact of subjective factors, a new breed of “hybrid choice” models have been developed; these models allow including not only tangible attributes but also more intangible elements associated with users’ perceptions and attitudes, through latent variables [24]. Moreover, it has been shown that introducing latent variables (LV) helps to improve the choice model fit [25, 26].

Using the logged odds (logit) of being a regular user or not, separate models for the factors with and without latent variables are estimated. The logged odds of pedestrian behaviors are modeled as a function of various personality and traffic variables.

## 5. Model and Estimation

*5.1. Correlation Analysis.* Using Kendall’s tau-b method, we found that there is no internal consistency among the factors. So the correlation analysis between each contributing factor and pedestrian crossing behavior is presented.

(1) *Demographic and Socioeconomic Characteristics.* For the gender factor,  $\chi^2$  is 7.934, and  $P = 0.046 < 0.05$ . The  $P$  values of age ( $F = 6.541$ ) and income ( $\chi^2 = 21.467$ ) factors are also smaller than 0.05. Furthermore, the  $P$  values of career and education factors are larger than 0.05. Thus, gender, age, and income are significant factors.

(2) *Family Characteristics.* As the  $P$  value is lower than 0.05, marital status factor is related to crossing behavior choices. In terms of compliance rate, there is no significant difference between the married and unmarried pedestrians. However, there are differences in different violating behaviors. Having children also is a significant factor.

(3) *Trip Characteristics.* With the prevalence of private cars, having a driver license has become closely associated with the travelers’ behaviors. Pedestrians who have a driver license are significantly higher in compliance rate (58.48%) than the pedestrians without a driver license (53.37%). Trip purpose is also a significant factor.

(4) *Attributes of the Facilities.* Among all the facilities variables, the crosswalk length, green time, and red time influence

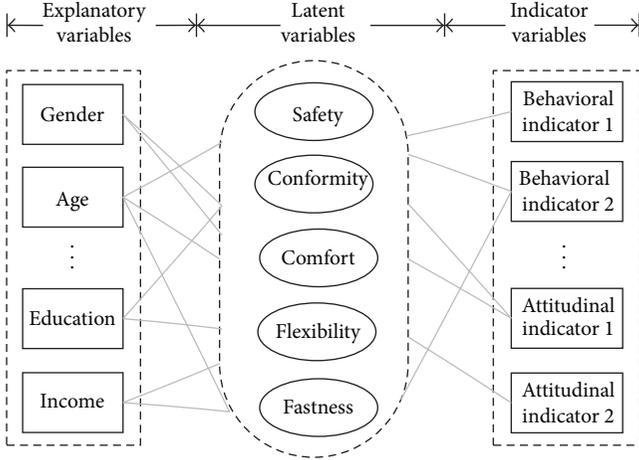


FIGURE 1: Latent variables model structure.

the crossing behaviors significantly, where the  $P$  values are 0.024, 0.042, and 0.006, respectively. But the crosswalk width ( $P = 0.71$ ) does not have significant effect.

(5) *Traffic Condition*. Among all the traffic condition variables, accompanied or alone, pedestrians waiting at the curb, the crossing pedestrians, the presence of oncoming cars and arrival time are significant.

5.2. *The Latent Model*. The science of behavior says that explanatory variables, such as gender, age, and income, result in a person's subjective preferences (latent variables), which could be reflected by the individual's behaviors and attitudes (indicator variables). The latent variable model describes the relationships between the latent variables and their indicators and causes. Figure 1 shows the relationships among the three types of variables.

Latent variables are unobserved attributes, which represent the pedestrians' subjective preferences in crossing behavior choice. In this study, the latent variables are safety ( $\eta_{\text{safe}}$ ), conformity ( $\eta_{\text{conf}}$ ), comfort ( $\eta_{\text{comf}}$ ), flexibility ( $\eta_{\text{flex}}$ ), and fastness ( $\eta_{\text{fast}}$ ). Latent variables are modeled in two aspects.

On one hand, the explanatory variables are used to construct latent variables as (1). We assume that the differences in pedestrian's attitude and preference are caused by demographic characteristics such as gender, age, and income:

$$\eta = \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_p x_p + \zeta, \quad (1)$$

where  $\eta$  is the latent variable and  $x_1, x_2, \dots, x_p$  are explanatory variables. And in this study, the explanatory variables include gender ( $x_{\text{gend}}$ ), age ( $x_{\text{age}}$ ), education ( $x_{\text{educ}}$ ), and income ( $x_{\text{inco}}$ ).  $\gamma_1, \gamma_2, \dots, \gamma_p$  ( $\gamma$ ) are the parameters to be estimated, and  $\zeta$  is an error term.

On the other hand, indicator variables can be represented by latent variables as seen in (2). Indicator variables are exogenous. We apply two different methods when constructing the latent variables. To construct the safety and conformity latent variables, we use behavioral indicator variables, and for

the comfort, flexibility, and fastness latent variables, we use attitudinal indicator variables. The questionnaire is designed as 5-point Likert Scale to describe the feelings of indicator variables:

$$\begin{aligned} y_1 &= \lambda_1 \eta + \varepsilon_1, \\ y_2 &= \lambda_2 \eta + \varepsilon_2, \dots, y_q = \lambda_q \eta + \varepsilon_q, \end{aligned} \quad (2)$$

where  $y_1, y_2, \dots, y_q$  are indicator variables,  $\lambda_1, \lambda_2, \dots, \lambda_q$  ( $\lambda$ ) are parameters, and  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_q$  are error terms. In this study, we have 16 indicator variables ( $y_1 - y_{16}$ ).

To build the model, gender variable is valued as 0 or 1, representing females or males. The age, education, and income variables are valued according to their levels. LISREL software is used to perform the latent variables modeling. Estimation results of  $\lambda$  and  $\gamma$  are shown in Tables 3 and 4, respectively.

Two aspects are used to evaluate the merits of the model. First, various fit indices are applied to the overall analysis of the model. The main evaluation indices as  $\chi^2/df$ , root mean square error of approximation (RMSEA), comparative fit index (CFI), and goodness-of-fit index (GFI) are 2.31, 0.016, 0.94, and 0.91, respectively; all the indices are in the acceptable range. Therefore, this model is viable.

Second, the significance levels of the parameters are examined and, thus, their meanings and suitability are evaluated. Results of the  $t$ -statistic output by LISREL are used to examine the significant level of the parameters. As shown in Table 3, the  $t$  values of all the parameters for  $\lambda$  are larger than 1.96, which means that all the parameters are significant at the 95% confidence level and, thus, the assumption is valid. However, only part of the  $t$  values of the  $\gamma$  parameters is larger than 1.96, and the others are lower than 1.96 (as shown in Table 4) and, therefore, should be eliminated.

Some correlations between latent variables and explanatory variables can be shown in Table 4. Pedestrians' preference for safety is positive with age and the level of income and negative with gender; conformity is negative with income, gender, and education; comfort is positive with income, age and education; flexibility is positive with gender while negative with age; finally, fastness is positively correlated with income and gender but negative with age.

### 5.3. The Multinomial Logit Model

5.3.1. *Model Variables*. After screening out the factors using the correlation analysis, the significant variables which are included in the model are classified as dichotomous data, ordered categorical variable, or continuous variables.

When used as the independent variables, dichotomous variables are usually valued as 0 and 1, respectively. Dichotomous variables in this study include gender, having a driver license or not, trip purpose, marital status, with or without children, accompanied or alone, and the oncoming cars. Ordered categorical variables could be analyzed using a dummy variable. The only ordered categorical variable is the level of income. There are totally four levels of the income, which are defined as three dummy variables. Generally, the higher level is chosen as the reference. Continuous variables

TABLE 3: The  $\lambda$  matrix of factor loadings.

Indicator variables	Latent variables									
	$\eta_{\text{safe}}$		$\eta_{\text{conf}}$		$\eta_{\text{comf}}$		$\eta_{\text{flex}}$		$\eta_{\text{fast}}$	
	Coef.	<i>t</i> -stat.								
$y_1$	1		—	—	—	—	—	—	—	—
$y_2$	0.69	9.85	—	—	—	—	—	—	—	—
$y_3$	0.97	12.97	—	—	—	—	—	—	—	—
$y_4$	1.39	14.12	—	—	—	—	—	—	—	—
$y_5$	—	—	1		—	—	—	—	—	—
$y_6$	—	—	1.58	13.76	—	—	—	—	—	—
$y_7$	—	—	0.43	5.34	—	—	—	—	—	—
$y_8$	—	—	—	—	1	—	—	—	—	—
$y_9$	—	—	—	—	1.11	11.42	—	—	—	—
$y_{10}$	—	—	—	—	0.88	10.80	—	—	—	—
$y_{11}$	—	—	—	—	—	—	1	—	—	—
$y_{12}$	—	—	—	—	—	—	0.85	9.53	—	—
$y_{13}$	—	—	—	—	—	—	0.38	4.11	—	—
$y_{14}$	—	—	—	—	—	—	—	—	1	—
$y_{15}$	—	—	—	—	—	—	—	—	1.53	13.15
$y_{16}$	—	—	—	—	—	—	—	—	0.72	8.59

TABLE 4: The  $\gamma$  matrix.

Latent variables	Explanatory variables									
	$x_{\text{gend}}$		$x_{\text{age}}$		$x_{\text{educ}}$		$x_{\text{inco}}$			
	Coef.	<i>t</i> -stat.	Coef.	<i>t</i> -stat.	Coef.	<i>t</i> -stat.	Coef.	<i>t</i> -stat.		
$\eta_{\text{safe}}$	-0.037	-2.84	0.231	14.57	-0.013	-1.25	0.075	5.23		
$\eta_{\text{conf}}$	-0.102	-7.42	0.009	1.04	-0.047	-2.38	-0.112	-7.89		
$\eta_{\text{comf}}$	-0.016	-0.34	0.092	8.68	0.032	2.54	0.186	11.80		
$\eta_{\text{flex}}$	0.044	2.01	-0.065	-5.76	0.15	1.31	-0.028	-1.87		
$\eta_{\text{fast}}$	0.042	2.95	-0.083	-7.51	0.07	1.50	0.191	11.40		

are typically put directly in the model, which include age, crosswalk length, the number of pedestrians waiting at the curb, the number of crossing pedestrians, and arrival time.

5.3.2. *MNL Model Estimation.* Based on the analysis from Section 3.2, the model contains four alternative parts (regular users, late starters, sneakers, and partial sneakers). The all-green type is chosen as the reference; thus, the logit model of the  $i$ th behavior could be written as

$$\ln\left(\frac{P_i}{P_1}\right) = \alpha^i + \sum_{k=1}^K \beta_k^i x_k^i. \quad (3)$$

Among them,  $P_1, P_2, P_3,$  and  $P_4$  are the probabilities of choosing to be a regular user, late starter, sneaker, and partial sneaker, respectively.  $\alpha^i$  is the constant, and  $\beta_k^i$  is the parameter of the explanatory variable  $x_k^i$ . Also,  $x_k^i$  indicates the  $k$ th explanatory variable when choosing the  $i$ th crossing behavior, such as gender, age, income1, and income2 in the model.

Using Biogeme software, models with latent variables are created, and the estimation results are shown in Table 5.

5.3.3. *Discussion.* From the estimation results of the parameters, for  $\ln(P_2/P_1)$  (the probability ratio of being late starters and regular users), the variable “arrival time” is the largest parameter value (2.04). It is seen that those who arrive during the last few seconds of the green time tend to be late starters, not the regular users. The second largest variable is “oncoming car;” since when there is no oncoming traffic, pedestrians are more likely to take the risk of crossing the street. Moreover, variables such as “length” and “green time” indicate that pedestrians are more likely to fail in crossing the street during the green time with a longer crosswalk and a shorter green time. Therefore, these factors significantly affect the crossing behaviors of this type primarily due to the traffic conditions and the road facilities. Furthermore, as for latent variables, the stronger subjective inclination for fastness, safety, and flexibility is, the more likely the pedestrian chooses to be a late starter.

For  $\ln(P_3/P_1)$  (the probability ratio of being sneakers and regular users), the variable “gender” has the largest parameter value (-3.75), showing that gender affects most significantly pedestrians’ choice of crossing on red, and the two are negatively correlated. Other significant factors include “oncoming car,” “trip purpose,” and “countdown.” For the latent variables,

TABLE 5: MNL model estimation results (with latent variables).

Independent variables	$\ln(P_2/P_1)$ late starter		$\ln(P_3/P_1)$ sneaker		$\ln(P_4/P_1)$ partial sneaker	
	Coef.	<i>t</i> -stat.	Coef.	<i>t</i> -stat.	Coef.	<i>t</i> -stat.
Gender	0.67	2.45	-3.75	-4.36	1.49	2.94
Age	1.37	3.05	—	—	-1.78	-3.96
Income1	—	—	—	—	—	—
Income2	—	—	-0.95	-2.87	1.25	3.53
Income3	1.25	3.16	-1.24	-2.05	—	—
License	-1.52	-6.48	-0.57	-3.91	1.05	3.97
Purpose	1.42	3.44	1.86	3.12	—	—
Married	0.58	2.37	0.63	2.05	-0.48	-2.72
Child	—	—	0.89	3.27	-0.75	-2.46
Length	1.75	3.23	-0.73	-1.98	-1.04	-2.95
Green time	-1.56	-4.82	—	—	-0.62	-2.03
Red time	-1.11	-2.75	0.52	4.35	—	—
Countdown	1.49	4.76	-1.75	-3.02	—	—
Accompanied	0.71	2.23	-1.53	-6.16	0.51	5.14
Waiting men	-0.26	-4.07	-0.53	-5.07	—	—
Crossing men	0.61	2.72	1.15	6.32	0.83	3.42
Oncoming car	-1.87	-3.75	-2.43	-2.78	-1.32	-3.05
Arrival time	2.04	2.45	-1.45	-3.71	0.57	2.34
$\eta_{\text{safe}}$	-0.45	-2.69	-1.01	-3.21	—	—
$\eta_{\text{conf}}$	—	—	1.37	2.98	0.92	2.25
$\eta_{\text{comf}}$	—	—	—	—	0.86	3.87
$\eta_{\text{flex}}$	0.67	2.96	—	—	—	—
$\eta_{\text{fast}}$	1.43	2.35	1.14	2.16	0.62	1.99
Constant	-3.75	-4.28	-1.69	3.38	4.37	3.06

Note: goodness-of-fit; McFadden  $\rho^2 = 0.426$ .

conformity, fastness, and safety greatly affect the choice of the pedestrian to cross on red. The influential level of safety is smaller than that of fastness, showing that pedestrians would choose a greater walking speed and efficiency at the price of safety.

Finally, for  $\ln(P_4/P_1)$  (the probability ratio of being partial sneakers and regular users), the variable “age,” with the parameter value of  $-1.78$ , is the most significant affecting factor. Other factors such as “gender,” “oncoming car,” and “income” also greatly affect pedestrians’ behavior choice of being partial sneakers. Apart from that, pedestrians’ subjective preferences for conformity, flexibility, and fastness also affect their behavior choosing. Thus, those who have a high requirement for these three features are more likely to be partial sneakers. In other words, when many pedestrians cross before green signal and the comfort of the crossing environment is poor, pedestrians are more likely to cross, which is especially true for those with urgent purposes.

It can be seen from the above that, on different behaviors, the effects of socioeconomic and traffic factors are different largely. However, there are still some similarities. For example, “oncoming car” is a factor that is in the top three great effect factors on the three behaviors. So, traffic condition is an important factor for pedestrians considering violate or not.

## 6. Conclusions

**6.1. Final Analysis.** To acquire more detailed information about pedestrian violations, this study examines pedestrians’ behavior characteristics and the affecting factors at signalized intersections. Information was obtained by field studies observation, where 1878 pedestrians were surveyed at 16 intersections in Nanjing, China. Conclusions are drawn based upon the correlation analysis and MNL model.

The main findings are as follows: (1) the presence of oncoming cars, crosswalk length, and green time are the most significant factors when pedestrians choose to be late starters; (2) gender, oncoming cars, and trip purpose are the most significant factors when pedestrians choose to be sneakers; and (3) age, gender, and oncoming cars are the most significant factors when pedestrians choose to be partial sneakers.

### 6.2. Recommendations

**6.2.1. Facility Design.** As is found from the presented analysis, the traffic conditions and road facilities significantly affect the crossing behaviors of pedestrians. To strengthen pedestrians’ awareness of obeying the rules, the facilities should be carefully designed for pedestrian crossings.

Firstly, pedestrians' needs should be considered in the design of crosswalks. For example, the shortest route for a crosswalk must be chosen to save energy consumption and reduce any mental obstacles.

Secondly, signal countdown devices should be installed to offer the information better. Therefore, pedestrians could make more effective decisions, which may decrease the illegal behaviors.

Finally, the design of green time for pedestrians should also be improved. In China, the crossing pedestrians and the vehicles moving in the same direction are allowed to pass simultaneously in most cases. Although this is simple, the signal facilities for the pedestrians are ineffective. When designing the signal, neighboring land use and the results from traffic survey should also be combined to judge the main pedestrian groups at this site. Moreover, the designed speed for the crossing pedestrians should be based on the characteristics of their gender, age, and the attributes of the facilities (e.g., with or without a countdown timer) at the intersection. That is, the shortest crossing time for the pedestrians could be determined methodically.

**6.3. Safety Education.** In this study, it was shown that the following groups are more likely to break the rules at a crosswalk: young and middle-aged people, females, and low-income people. Therefore, various methods for safety education should be applied to best suit the pedestrians in these groups. For example, female pedestrians should be told more about the rules, as some of them are not familiar with traffic rules in China. In addition, social psychology should be considered to give effective propagation and education to the public. Also, computer and networks technology should be harnessed so as to open up new areas for traffic propaganda and public education.

## Acknowledgments

This research is supported by the National Natural Science Foundation of China (51308298) and Project of Nanjing University of Science and Technology.

## References

- [1] TABC, *Statistics of Road Traffic Accidents in China (2009)*, Traffic Administration Bureau of China State Security Ministry, Beijing, China, 2010.
- [2] National Highway Traffic Safety Administration, "Traffic safety facts 2006," Report DOT HS 811 402, National Highway Traffic Safety Administration, Washington, DC, USA, 2009.
- [3] TABC, *Statistics of Road Traffic Accidents in PR in China (2008)*, Traffic Administration Bureau of China, Ministry of Public Security, Beijing, China, 2009.
- [4] J. Yang, W. Deng, J. Wang, Q. Li, and Z. Wang, "Modeling pedestrians' road crossing behavior in traffic system micro-simulation in China," *Transportation Research A*, vol. 40, no. 3, pp. 280–290, 2006.
- [5] G. Ren, Z. Zhou, W. Wang, Y. Zhang, and W. Wang, "Crossing behaviors of pedestrians at signalized intersections: observational study and survey in China," *Transportation Research Record*, no. 2264, pp. 65–73, 2011.
- [6] J. Bergeron, J. P. Thouez, H. Bélanger, R. Bourbeau, D. Lord, and A. Rannou, "Study on conflicts among pedestrians and drivers," in *Proceedings of the 9th PRI World Congress*, Madrid, Spain, 2002.
- [7] O. Keegan and M. O'Mahony, "Modifying pedestrian behaviour," *Transportation Research A*, vol. 37, no. 10, pp. 889–901, 2003.
- [8] X. Chu, M. Guttenplan, and M. R. Baltes, "Why people cross where they do: the role of street environment," in *Transportation Research Record: Journal of Transportation Research Board*, vol. 1878, pp. 3–10, Transportation Research Board of National Academics, Washington, DC, USA, 2004.
- [9] G. Tiwari, S. Bangdiwala, A. Saraswat, and S. Gaurav, "Survival analysis: pedestrian risk exposure at signalized intersections," *Transportation Research F*, vol. 10, no. 2, pp. 77–89, 2007.
- [10] T. Rosenbloom, D. Nemrodov, and H. Barkan, "For heaven's sake follow the rules: pedestrians' behavior in an ultra-orthodox and a non-orthodox city," *Transportation Research F*, vol. 7, no. 6, pp. 395–404, 2004.
- [11] M.-A. Granié, "Gender differences in preschool children's declared and behavioral compliance with pedestrian rules," *Transportation Research F*, vol. 10, no. 5, pp. 371–382, 2007.
- [12] C. Holland and R. Hill, "The effect of age, gender and driver status on pedestrians' intentions to cross the road in risky situations," *Accident Analysis and Prevention*, vol. 39, no. 2, pp. 224–237, 2007.
- [13] E. Moyano Díaz, "Theory of planned behavior and pedestrians' intentions to violate traffic regulations," *Transportation Research F*, vol. 5, no. 3, pp. 169–175, 2002.
- [14] I. M. Bernhoft and G. Carstensen, "Preferences and behaviour of pedestrians and cyclists by age and gender," *Transportation Research F*, vol. 11, no. 2, pp. 83–95, 2008.
- [15] M. M. Hamed, "Analysis of pedestrians' behavior at pedestrian crossings," *Safety Science*, vol. 38, no. 1, pp. 63–82, 2001.
- [16] N. Guéguen and N. Pichot, "The influence of status on pedestrians' failure to observe a road-safety rule," *Journal of Social Psychology*, vol. 141, no. 3, pp. 413–415, 2001.
- [17] D. C. Schwebel, J. Severson, K. K. Ball, and M. Rizzo, "Individual difference factors in risky driving: the roles of anger/hostility, conscientiousness, and sensation-seeking," *Accident Analysis and Prevention*, vol. 38, no. 4, pp. 801–810, 2006.
- [18] D. A. Santor, D. Messervey, and V. Kusumakar, "Measuring peer pressure, popularity, and conformity in adolescent boys and girls: predicting school performance, sexual attitudes, and substance abuse," *Journal of Youth and Adolescence*, vol. 29, no. 2, pp. 163–182, 2000.
- [19] T. Rosenbloom, "Crossing at a red light: Behaviour of individuals and groups," *Transportation Research F*, vol. 12, no. 5, pp. 389–394, 2009.
- [20] R. Zhou, W. J. Horrey, and R. Yu, "The effect of conformity tendency on pedestrians' road-crossing intentions in China: an application of the theory of planned behavior," *Accident Analysis and Prevention*, vol. 41, no. 3, pp. 491–497, 2009.
- [21] V. P. Sisiopiku and D. Akin, "Pedestrian behaviors at and perceptions towards various pedestrian facilities: an examination based on observation and survey data," *Transportation Research F*, vol. 6, no. 4, pp. 249–274, 2003.
- [22] M. F. Yáñez, S. Raveau, and J. D. D. Ortúzar, "Inclusion of latent variables in Mixed Logit models: modelling and forecasting," *Transportation Research A*, vol. 44, no. 9, pp. 744–753, 2010.

- [23] T. A. Domencich and D. McFadden, *Urban Travel Demand: A Behavioral Analysis*, North-Holland, Amsterdam, The Netherlands, 1975.
- [24] M. E. Ben-Akiva, J. L. Walker, A. T. Bernardino, D. A. Gopinath, and T. Morikawa, "Integration of choice and latent variable models," in *Perpetual Motion: Travel Behavior Research Opportunities and Challenges*, H. S. Mahmassani, Ed., Pergamon Press, Amsterdam, The Netherlands, 2002.
- [25] K. Ashok, W. R. Dillon, and S. Yuan, "Extending discrete choice models to incorporate attitudinal and other latent variables," *Journal of Marketing Research*, vol. 39, no. 1, pp. 31–46, 2002.
- [26] M. Vredin Johansson, T. Heldt, and P. Johansson, "The effects of attitudes and personality traits on mode choice," *Transportation Research A*, vol. 40, no. 6, pp. 507–525, 2006.

## Research Article

# Pattern Analysis of Driver's "Pressure-State-Response" in Traffic Congestion

Weiwei Qi,<sup>1</sup> Yulong Pei,<sup>2</sup> Mo Song,<sup>3</sup> and Yiming Bie<sup>1</sup>

<sup>1</sup> School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

<sup>2</sup> College of Traffic, Northeast Forestry University, Harbin 150040, China

<sup>3</sup> Department of Automobile Service Engineering, Zhejiang Traffic Technician College, Jinhua 321000, China

Correspondence should be addressed to Weiwei Qi; [qwwhit@163.com](mailto:qwwhit@163.com)

Received 25 September 2013; Revised 4 November 2013; Accepted 10 November 2013

Academic Editor: Huimin Niu

Copyright © 2013 Weiwei Qi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traffic congestion, which has a direct impact on the driver's mood and action, has become a serious problem in rush hours in most cities of China. Currently, the study about driver's mood and action in traffic congestion is scarce, so it is necessary to work on the relationship among driver's mood and action and traffic congestion. And the PSR (pressure-state-response) framework is established to describe that relationship. Here, PSR framework is composed of a three-level logical structure, which is composed of traffic congestion environment, drivers' physiology change, and drivers' behavior change. Based on the PSR framework, various styles of drivers have been chosen to drive on the congested roads, and then traffic stream state, drivers' physiology, and behavior characters have been measured via the appropriate equipment. Further, driver's visual characteristics and lane changing characteristics are analyzed to determine the parameters of PSR framework. According to the PSR framework, the changing law of drivers' characteristics in traffic congestion has been obtained to offer necessary logical space and systematic framework for traffic congestion management.

## 1. Introduction

Traffic congestion has become a peculiar phenomenon in rush hour of big city, and the rapid increasing number of automobiles and comparable insufficiency of transportation facilities are the direct reason [1]. So, scholars usually research the causes, formation mechanism, and mitigation strategies of traffic congestion from the perspective of traffic supply and traffic demand. Arnott [2] established a bathtub model of downtown rush-hour traffic congestion to perfect the standard economic models of traffic congestion. Tsekeris and Geroliminis [3] analyzed the relationship between land use and traffic congestion by employing the macroscopic fundamental diagram, which constitutes robust second-best optimal strategies that can further reduce congestion externalities. Traffic congestion prediction plays an important role in route guidance and traffic management [4], and many traffic congestion prediction models have been proposed by scholars, such as the nearest neighbor method [5], the ARIMA

(autoregressive integrated moving average) model [6] and the vector ARMA (autoregressive moving average) model [7].

Traffic congestion has brought huge economic losses and adverse impact on the driver's mood [8]. Traffic congestion increases the drivers' physiological pressures and the burdens of visual cognition [9], which leads to risky driving behavior [10]. So, "perception-judgment-decision" process reflects the formation mechanism for the driving behavior in rush hour of urban road.

According to statistical data, 70% of the driver's perception information is based on visual system, and many studies have recorded and analyzed drivers' eye movements. Lansdown [11] completed a study in which visual allocation and verbal reports were recorded to determine individual differences in drivers conducting in-vehicle tasks. Underwood et al. [12] argued that it was common sense that a driver must look at the appropriate locations in a traffic scene in order to gain information about risks and potential risks in the scene. Benedetto et al. [13] obtained more short blinks that occur

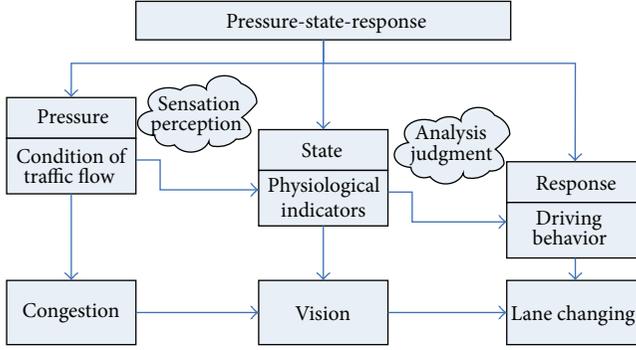


FIGURE 1: Model of driver's "pressure-state-response".

with an IVIS (in-vehicle information systems) interaction during driving, while more long blinks arise as time spent driving increases.

In various driving situations, drivers with higher trait tendencies are considered more likely to experience frequent and intense emotional states [14]. The influence of driving anger tendencies on the driver's behavior has not been clear. Anger-prone drivers have reported driving at higher speed and with less speed limit compliance as well as more near accidents [15, 16], less concentration, and reduced vehicular control while driving [14, 15].

So, traffic congestion would have a negative impact on the mood of the drivers on urban road [8], which will be reflected in the eye movements and driving behavior characteristics. This paper attempts to explore the regular pattern for "traffic congestion-eye movement-driving behavior."

## 2. "Pressure-State-Response" Model of Drivers in Traffic Congestion

In 1979, Rapport et al. [17, 18] proposed a PSR (pressure-state-response) model and concluded that humans and all other biota coevolve with their environments. Then, the PSR model was widely used in ecological security assessment, land quality assessment, evaluation of sustainable land use, and ecosystem health assessment [19, 20]. In the state of traffic congestion, the interaction between drivers and traffic environment can be described through an improved PSR model, as shown in Figure 1. So, the definition and structure for the improved PSR model are the research focus of this section. In order to explore the regular pattern for the improved PSR model (pressure-traffic congestion, state-eye movement, and response-lane changing), the frame model in traffic congestion can be regarded as a three-level logical structure, which can be explained as follows.

- (1) Pressure-traffic congestion: drivers start a travel with some purposes, and they want to drive under a desired speed, but the traffic congestion will reduce the speed of their vehicles. Daily driving, particularly in congestion condition, can be viewed as a frequent source of stress [21]. Therefore, the pressure can be measured by the degree of traffic congestion.

- (2) State-eye movement: it is the change of drivers' physiological indexes under the condition of traffic congestion, which can be measured by driver's visual characteristics, such as fixation points' distribution, fixation duration, average saccade speed, average saccade acceleration, blink duration, and blink rate.
- (3) Response-lane changing: drives' stress has been found to subsequently influence mood, thoughts, feelings, and behaviors [22]. The response is the driving action that drivers take under the specific physiological condition due to traffic congestion, and the response can be measured by driver's lane changing characteristics.

## 3. Source Pressure for Drivers in Traffic Congestion

Traffic congestion is the loss of travel time and running speed for drivers; thus, pressure coefficient  $\chi_{\text{press}}^{t_0-t_n}$  is defined to express the stress that drivers are subjected to in traffic congestion. Pressure coefficient  $\chi_{\text{press}}^{t_0-t_n}$ , which is the product of drivers' time and speed, is a cumulative index due to traffic congestion from  $t_0$  to  $t_n$  in rush hour. It can be shown as follows:

$$\chi_{\text{press}}^{t_0-t_n} = \frac{\left( \int_{t=t_0}^{t_n} V_{\text{off-peak}}(t) dt - \int_{t=t_0}^{t_n} V_{\text{rush}}(t) dt \right)}{\int_{t=t_0}^{t_n} V_{\text{off-peak}}(t) dt}, \quad (1)$$

where  $\chi_{\text{press}}^{t_0-t_n}$  is the pressure coefficient of drivers driving on route  $S$  from  $t_0$  to  $t_n$  during rush hour, km;  $V_{\text{rush}}(t)$  is the function of time, representing driving speed on route  $S$  during rush hour, km/h;  $V_{\text{off-peak}}(t)$  is the function of time, representing driving speed on route  $S$  during nonrush hour, km/h;  $t$  is the travel time of drivers, h;  $t_0$  is the departure time of drivers on route  $S$  during rush hour, h;  $t_n$  is the arrival time of drivers on route  $S$  during rush hour, h.

According to the definition of the pressure coefficient in traffic congestion, it can be simplified as follows:

$$\chi_{\text{press}}^{t_0-t_n} = \frac{[\bar{V}_{\text{off-peak}} \cdot (t_n - t_0) - \bar{V}_{\text{rush}} \cdot (t_n - t_0)]}{[\bar{V}_{\text{off-peak}} \cdot (t_n - t_0)]}, \quad (2)$$

$$\chi_{\text{press}}^{t_0-t_n} = \frac{(\bar{V}_{\text{off-peak}} - \bar{V}_{\text{rush}})}{\bar{V}_{\text{off-peak}}}, \quad (3)$$

where  $\bar{V}_{\text{rush}}$  is the average speed of vehicles on route  $S$  during rush hour, km/h;  $\bar{V}_{\text{off-peak}}$  is the average speed of vehicles on route  $S$  during nonrush hour, km/h.

Evaluation questionnaire on traffic congestion is a draft based on the definition of the pressure coefficient in traffic congestion, and the analyzing results of 1000 questionnaires are shown in Figure 2. The statistical results of 1000 questionnaires are fitted to normal distribution function ( $\mu = 0.5087$ ,  $\sigma = 0.1361$ , and  $R^2 = 0.9656$ ); therefore, pressure coefficient for the 85% cumulative frequency value is 0.65; thus  $\chi_{\text{press}}^{t_0-t_n} = 0.65$  can be treated as the threshold value to classify the road traffic volume state. When  $\chi_{\text{press}}^{t_0-t_n} < 0.65$ , traffic volume can

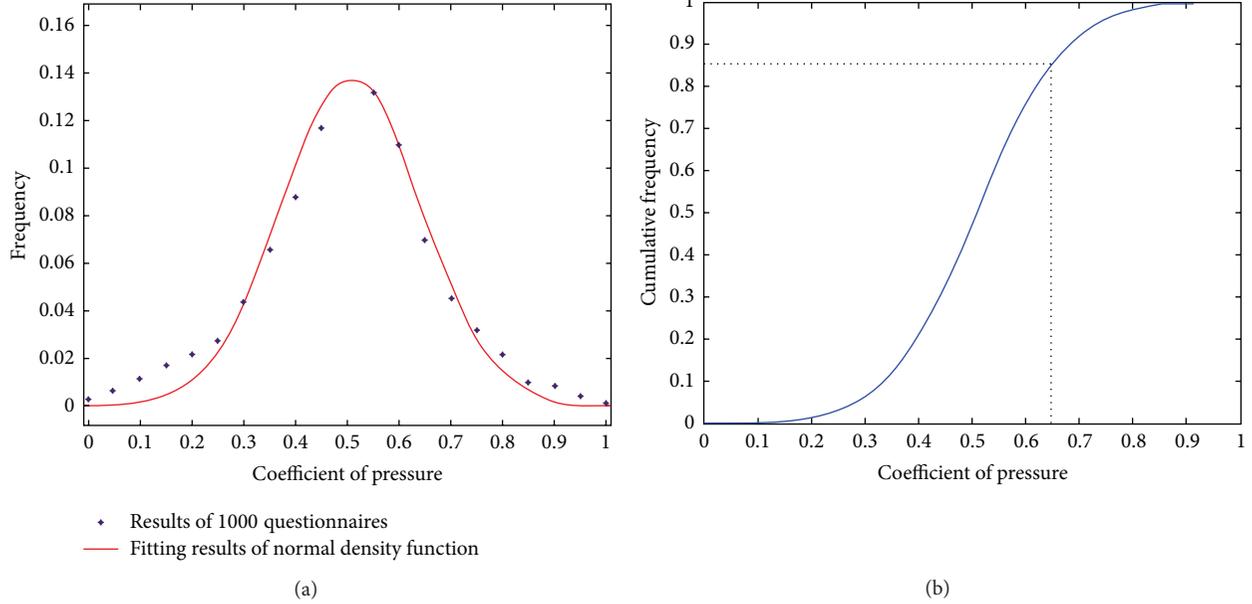


FIGURE 2: Evaluation results of pressure coefficient in traffic congestion.

be regarded as being in general state; while  $\chi_{\text{press}}^{t_0-t_n} > 0.65$ , it is in the congested state. Next, the drivers' visual characteristics and lane changing characteristics will be compared separately during the two states.

#### 4. Driver's Visual Characteristics in Traffic Congestion

**4.1. Experimental Design.** Eighty drivers (30 females, mean age = 33, min = 21, max = 50, and SD = 8; 50 males, mean age = 36, min = 21, max = 58, and SD = 11) were recruited and informed about the experiment's general purposes. All of them declared that they had valid Chinese driving licenses, with a mean of 9 years of driving experience. The dynamic visual characteristics of the drivers are selected as indicators of "state" on the basis for pressure source of traffic congestion, and the test scenarios of drivers' visual characteristics are shown in Figure 3. All participants were informed about the possibility of giving up (without any consequences) at any time if they did not feel comfortable during the experiment. All of them finished the process about data recording of their eye movements and driving performance in the general and congested state (that two states are distinguished via formula (3) and the threshold value of  $\chi_{\text{press}}^{t_0-t_n}$ ).

Whether the change of drivers' eye movements is obvious and what kind of characteristics drivers' vision may have in traffic congestion are studied based on the test about the eye movement characteristics of 80 drivers in general and congested state via an SMI iView X HED head-mounted monocular eye tracker. The test indicators (dependent variables) of participants' eye movements are shown in Table 1.

##### 4.2. Eye Fixation

**4.2.1. Fixation Points' Distribution.** Similarities and differences of fixation points in general and congested state are represented by the coordinates, and this indicates the spatial

TABLE 1: Indexes of drivers' visual characteristics in general and congested state.

No.	Types of eye movements	Indexes of eye movements
1	Fixation	Fixation points' distribution Fixation duration
2	Saccade	Average saccade speed Average saccade acceleration
3	Blink	Blink duration Blink rate

distribution of the fixation points.  $x$ -axis of the horizon plane is divided into 800 units and  $y$ -axis is divided into 600 units, as shown in Figure 4. According to Figure 4, in general state, fixation points of divers are mainly accumulated in an area where  $x$ -axis is from 150 to 500 and  $y$ -axis is from 300 to 550, while, in congested state, fixation points of divers are mainly accumulated in an area where  $x$ -axis is from 100 to 600 and  $y$ -axis is from 200 to 500. So, the space range of fixation points' distribution in congested state is 1.7 times that of in general state.

**4.2.2. Fixation Duration.** The fixation duration is an important indicator of the drivers' visual capacity allocation. And the results of statistical analysis are shown in Table 2. The mean value and standard deviation of the fixation duration are different in general and congested state. From Figures 5 and 6, it can be concluded that the fixation duration in general and congested state generally obeys log-normal distribution other than normal distribution.

##### 4.3. Eye Saccade

**4.3.1. Average Saccade Speed.** The statistical analysis results of drivers' average saccade speed are shown in Table 3. The mean



FIGURE 3: Test scenarios of driver's visual characteristics.

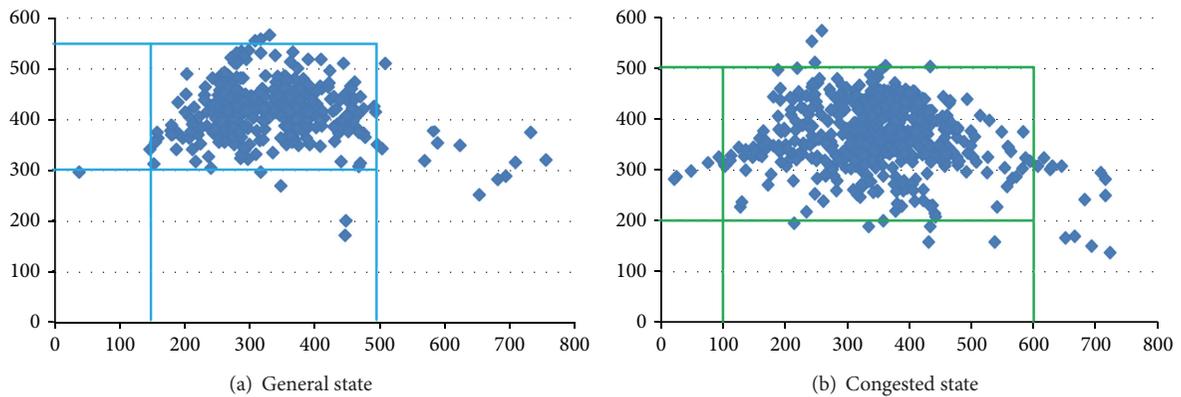


FIGURE 4: Planar distribution characteristics of driver's fixation points.

value and standard deviation of the average saccade speed are higher in congested state than those in general state. The average saccade speed in general and congested state generally obeys normal distribution other than log-normal distribution, as shown in Figures 7 and 8.

4.3.2. *Average Saccade Acceleration.* The statistical analysis results of drivers' average saccade acceleration are shown in Table 4. The mean value and standard deviation of the average saccade acceleration are different in general and congested state. The average saccade acceleration in general

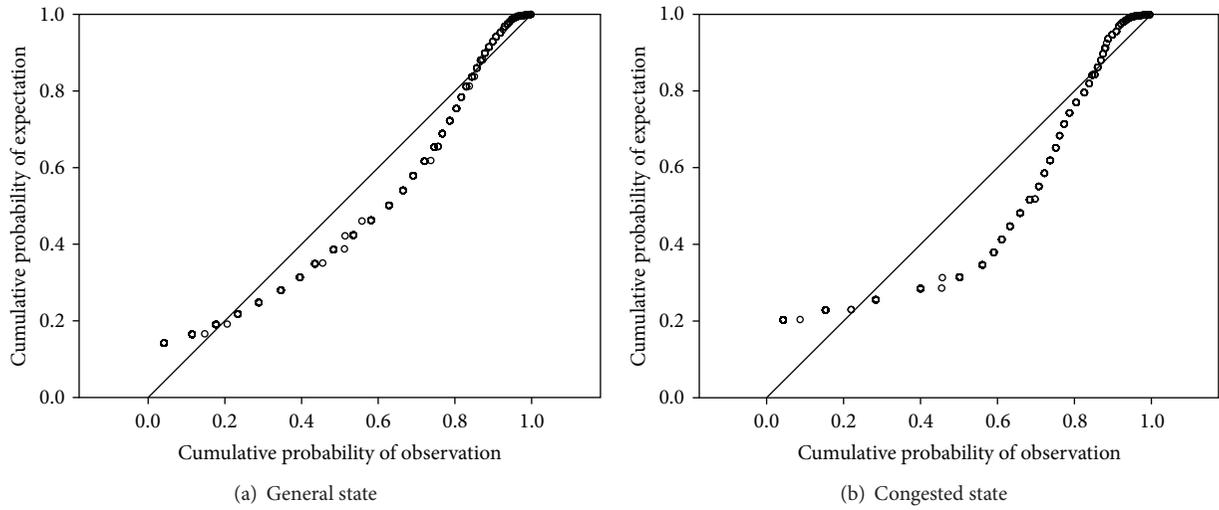


FIGURE 5: P-P figure of normal distribution fitting.

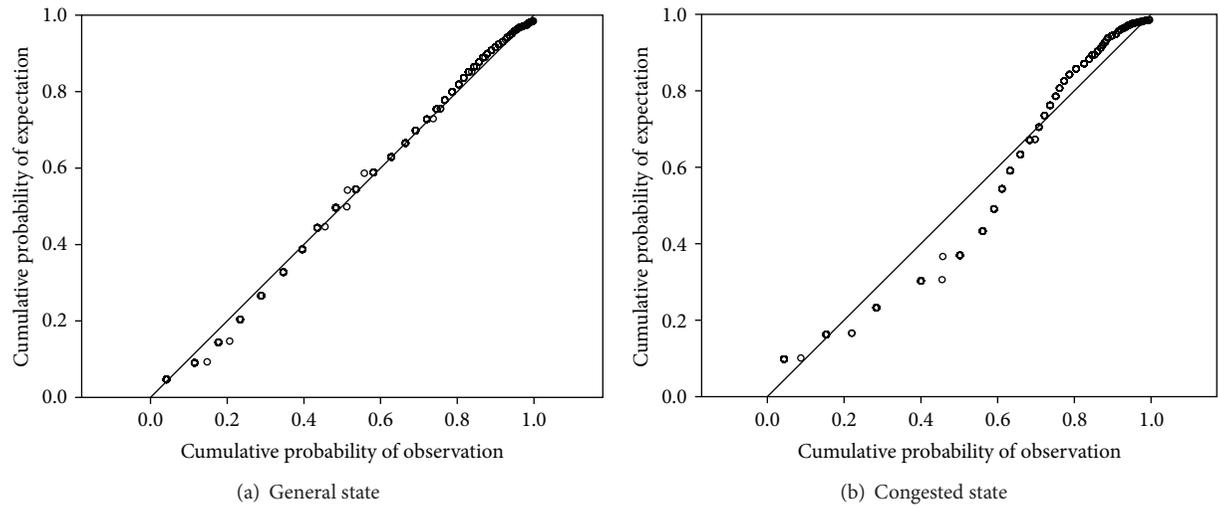


FIGURE 6: P-P figure of log-normal distribution fitting.

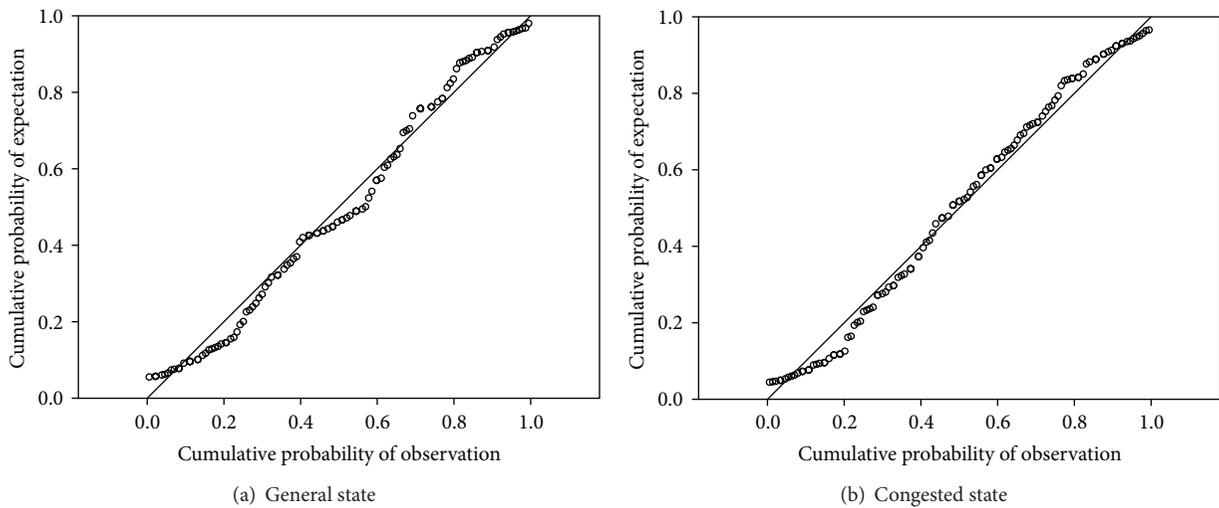


FIGURE 7: P-P figure of normal distribution fitting.

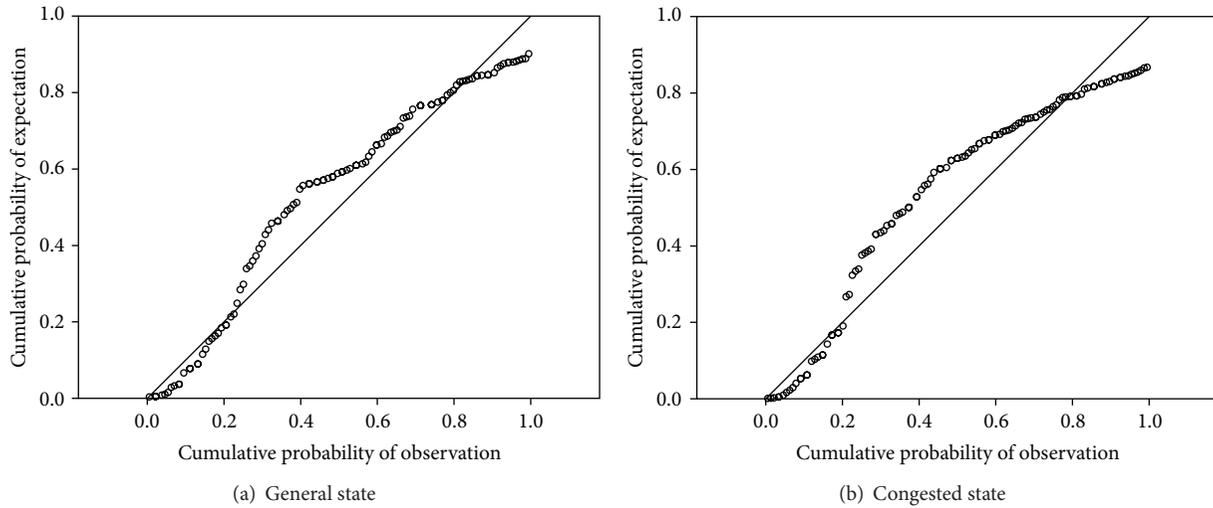


FIGURE 8: P-P figure of log-normal distribution fitting.

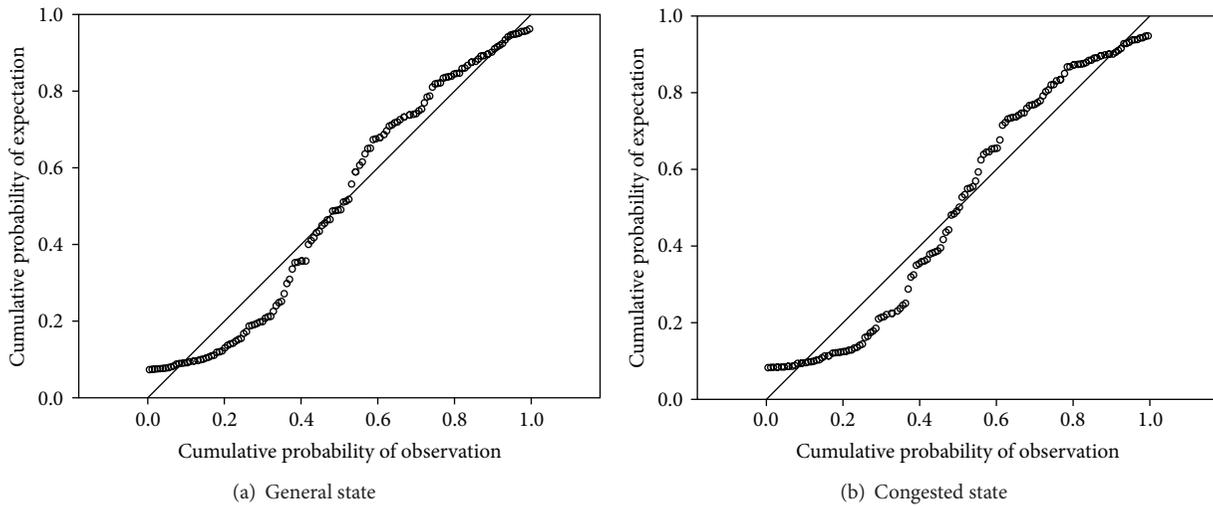


FIGURE 9: P-P figure of normal distribution fitting.

and congested state does not obey normal or log-normal distribution, as shown in Figures 9 and 10.

#### 4.4. Eye Blink

**4.4.1. Blink Duration.** The statistical analysis results of drivers' blink duration are shown in Table 5. From Figures 11 and 12, it can be concluded that the blink duration in general and congested state generally obeys log-normal distribution other than normal distribution.

**4.4.2. Blink Rate.** The statistical analysis results of drivers' blink rate are shown in Table 6. The mean value and standard deviation of the blink rate are different in general and congested state. From Figures 13 and 14, it can be concluded that the blink rate in general and congested state generally obeys normal distribution and log-normal distribution.

**4.5. Significance Test.** Based on the foregoing analysis, the indicators of drivers' visual characteristics could not meet the requirements (normal distribution and variance homogeneity) for parameter test. Therefore, the nonparametric test method (Mann-Whitney  $U$  test) is selected to test the difference significance for the indicators of drivers' visual characteristics in the general and congested state.

According to Table 7, the indicators of drivers' visual characteristics in the general and congested state are different in difference significance. Among them, the difference for the fixation duration and blink rate index is very significant ( $P \leq 0.01$ ); the difference for the average saccade speed and blink duration index is significant ( $P \leq 0.05$ ); but the difference for the average saccade acceleration index is not significant ( $P > 0.05$ ). So, our research results suggest that traffic congestion has a strong effect on driver's visual characteristics.

TABLE 2: Results of statistical analysis on the fixation duration.

No.	Sample	Average	Std.	Deviation	Kurtosis
1	General state	299	205	1.330	1.371
2	Congested state	271	229	1.550	1.486

TABLE 3: Results of statistical analysis on the average saccade speed.

No.	Sample	Average	Std.	Deviation	Kurtosis
1	General state	122	69	0.191	-0.972
2	Congested state	146	81	-0.033	-1.072

TABLE 4: Results of statistical analysis on the average saccade acceleration.

No.	Sample	Average	Std.	Deviation	Kurtosis
1	General state	8704	5958	0.050	-1.333
2	Congested state	9129	6510	0.043	-1.463

TABLE 5: Results of statistical analysis on the blink duration.

No.	Sample	Average	Std.	Deviation	Kurtosis
1	General state	212	97	0.736	-0.328
2	Congested state	202	94	0.856	-0.087

TABLE 6: Results of statistical analysis on the blink rate.

No.	Sample	Average	Std.	Deviation	Kurtosis
1	General state	0.25	0.05	0.341	-1.130
2	Congested state	0.22	0.05	-0.140	-0.935

TABLE 7: Difference significance for the indicators of drivers' visual characteristics.

No.	Index	General state		Congested state		Sig.
		Average	Std.	Average	Std.	
1	Fixation duration	299	205	271	229	0.000
2	Average saccade speed	122	69	146	81	0.019
3	Average saccade acceleration	8704	5958	9129	6510	0.553
4	Blink duration	212	97	202	94	0.023
5	Blink rate	0.25	0.05	0.22	0.05	0.006

## 5. Lane Changing Characteristics in Congested State

5.1. *Types and Statistical Characteristics of Risky Lane Changing.* While driving in crowded traffic flow, drivers are subjected to the pressure of low speed, and they would change lanes to improve driving condition and gain higher speed. With the increase of congestion pressure, many drivers would choose risky lane changing behaviors to obtain bigger space, which may generate traffic conflicts. Risky lane changing behaviors are divided into three types according to their different characteristics.

- (1) Lane changing directly: drivers directly import their vehicles into the target lane even if the headway does not meet the demands. It is a kind of risky lane changing behavior that would lead to serious conflicts.
- (2) Lane changing pressingly: drivers continue to squeeze their vehicles into the target lane even if the headway does not meet the demands, and once enough space is gained, they are imported into the target lane immediately. It is a kind of risky lane changing behavior that would lead to certain conflicts.
- (3) Lane changing selectively: drivers' vehicles and other vehicles in the target lane run in parallel, and drivers gradually turn their vehicles to the target lane even if the headway does not meet the demands. It is a kind of risky lane changing behavior that would lead to slight conflicts.

The statistical data of risky lane changing behaviors with different traffic parameters is as shown in Table 8. In Figures 15(a) and 15(b), we see that traffic volume and speed have the similar impact on the three types of risky lane changing behaviors. The percentage of lane changing selectively rises with the increase of traffic volume and speed, while the percentage of lane changing pressingly falls. Farther, the percentage of lane changing directly comes to its max value in the middle value of traffic volume and speed.

5.2. *Relationship between Frequency of Risky Lane Changing and Traffic Conflicts.* According to the selection of classification indexes for risky lane changing behaviors, influences for different types of risky lane changing behaviors on road safety are different, and the frequencies of risky lane changing and number of traffic conflicts per unit time at varied observation points are shown in Table 9.

The frequencies of lane changing directly, lane changing pressingly, and lane changing selectively are defined as independent variable of  $X_1$ ,  $X_2$ , and  $X_3$ ; the number of traffic conflicts is defined as a dependent variable of  $Y$ . A linear model of the traffic conflicts and different lane changing frequencies is established by regression analysis on data of Table 9. The model is shown in formula (4), and the value of residual and standard residual of the model which satisfies the requirements is shown in Table 10:

$$Y = 2.5X_1 + 1.4X_2 + 0.53X_3 + 13.29. \quad (4)$$

According to the absolute value of the influence coefficient for all kinds of risky lane changing behaviors on traffic conflicts, behavior of lane changing directly has the most influence on road safety, in which drivers changing lane directly for once can lead to 2.5 times of traffic conflicts in average. Accordingly, behavior of lane changing selectively has the least influence on road safety, in which drivers changing lane selectively for once can lead to 0.53 times of traffic conflicts in average.

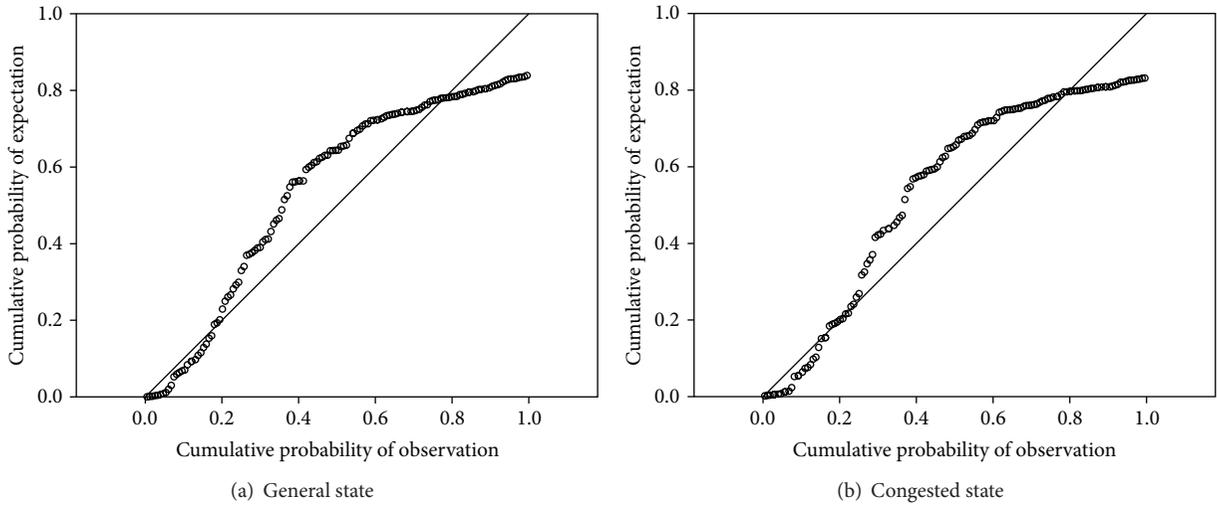


FIGURE 10: P-P figure of log-normal distribution fitting.

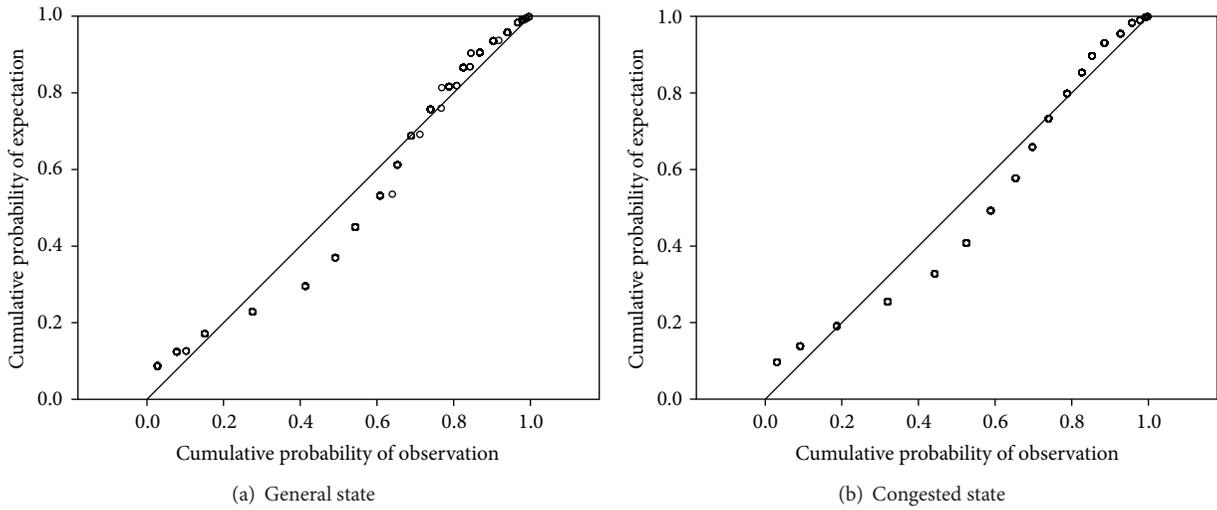


FIGURE 11: P-P figure of normal distribution fitting.

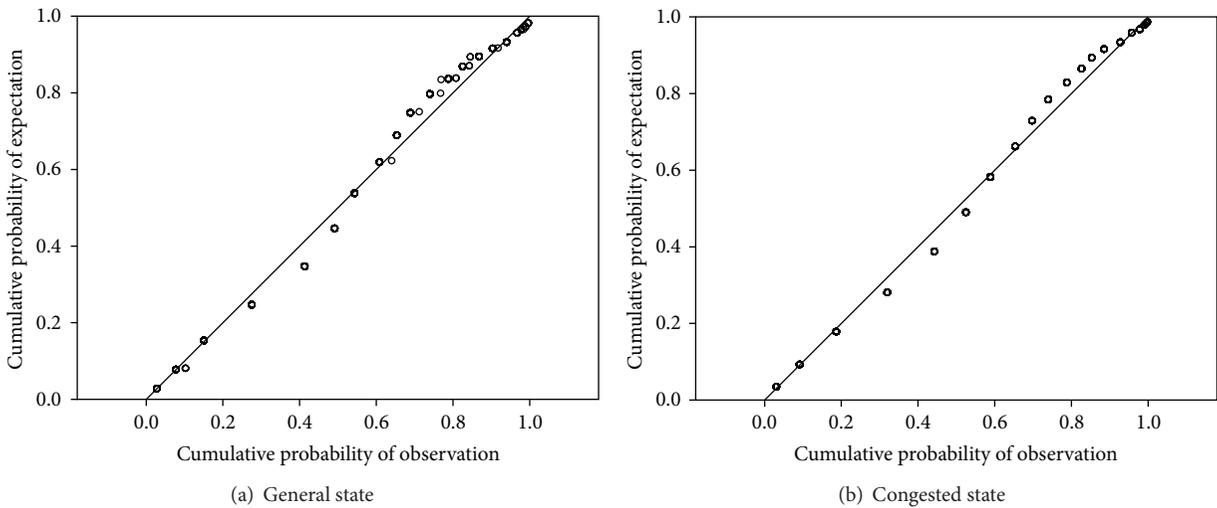


FIGURE 12: P-P figure of log-normal distribution fitting.

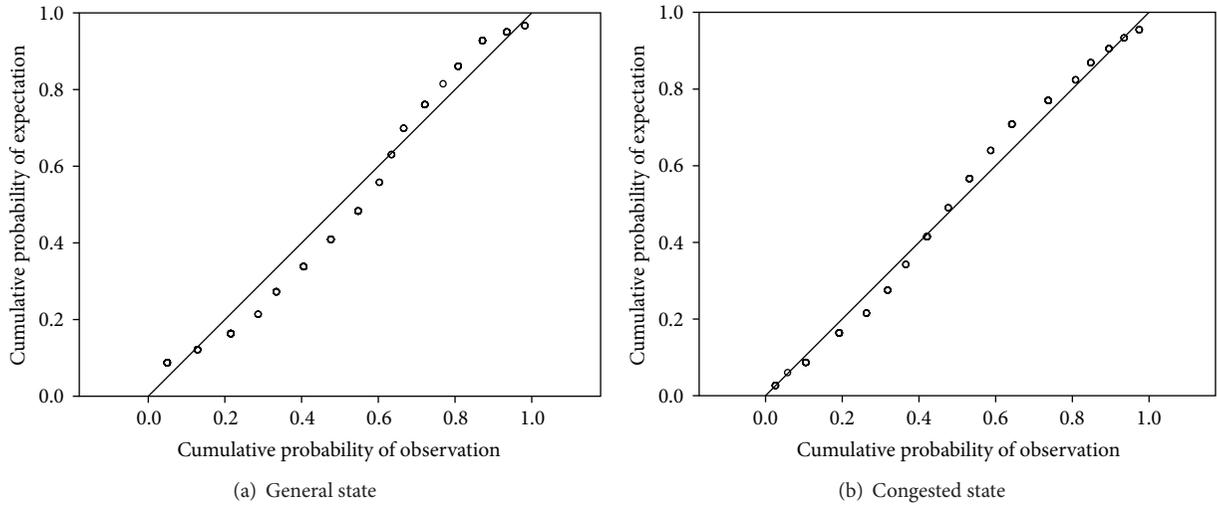


FIGURE 13: P-P figure of normal distribution fitting.

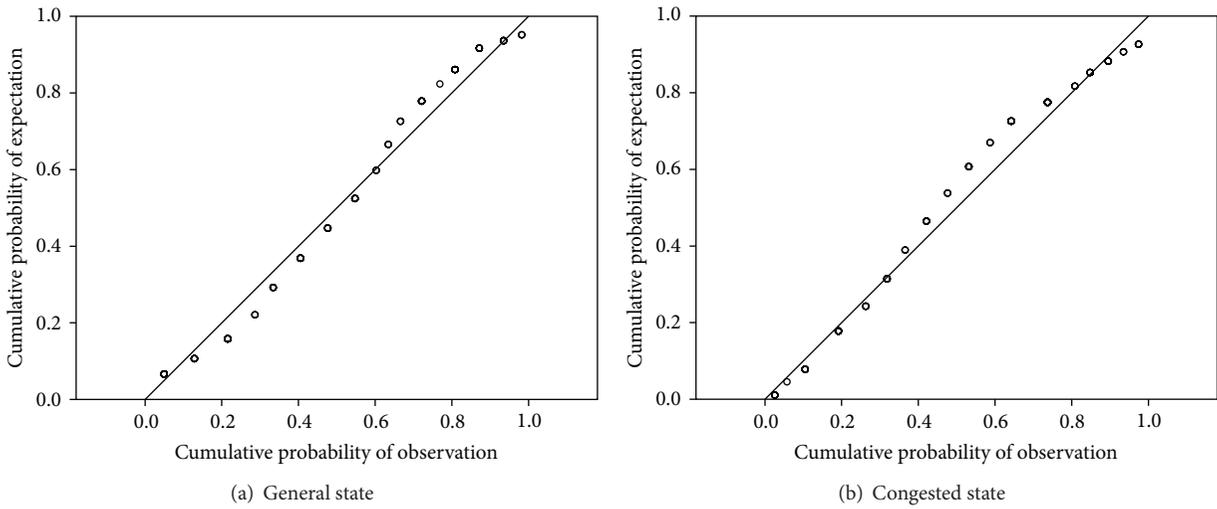


FIGURE 14: P-P figure of log-normal distribution fitting.

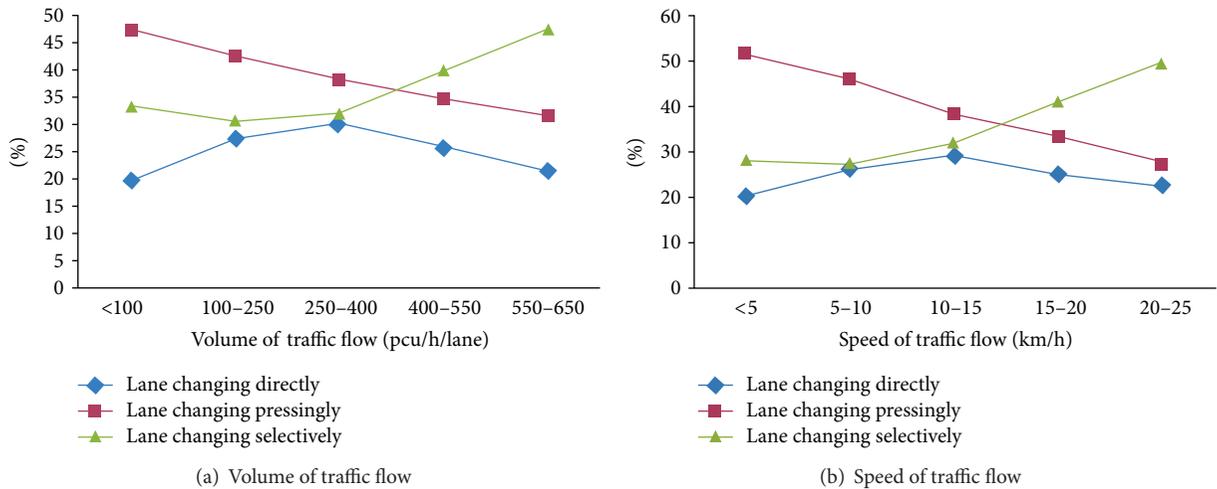


FIGURE 15: Average weight of different risky lane changing behaviors in typical cities of China.

TABLE 8: Statistical analysis of risky lane changing behaviors via different traffic parameters.

Type of risky lane changing		Lane changing directly (%)	Lane changing pressingly (%)	Lane changing selectively (%)
Volume of traffic flow (pcu/h/lane)	<100	19.71	47.18	33.11
	100–250	27.31	42.34	30.35
	250–400	30.29	37.92	31.79
	400–550	25.78	34.55	39.67
	550–650	21.38	31.41	47.21
Speed of traffic flow (km/h)	<5	20.34	51.41	28.25
	5–10	26.52	46.13	27.35
	10–15	29.39	38.43	32.18
	15–20	25.21	33.63	41.16
	20–25	22.78	27.83	49.39

TABLE 9: Statistical analyses for frequencies of risky lane changing and number of traffic conflicts.

No.	Frequencies of risky lane changing behaviors (times/h)			Traffic conflicts (times/h)
	Lane changing directly	Lane changing pressingly	Lane changing selectively	
1	35	23	19	143
2	22	18	27	108
3	18	29	15	107
4	16	22	23	96
5	28	16	25	119

TABLE 10: Value of residual and standard residual of formula (4).

Prediction Y	143.10	107.81	106.85	96.28	118.96
Residual	-0.0986	0.1899	0.1532	-0.2801	0.0357
Standard residual	-0.2555	0.4920	0.3968	-0.7257	0.0924

## 6. Conclusion

Drivers' physical and mental health in traffic congestion should attract more attention from traffic engineer and government administration. So, the impact of traffic congestion on drivers' eye movement and lane changing behavior has been quantified via the PSR model in this paper. In addition, the pressure coefficient has been defined to reflect drivers' feelings about the degree of traffic congestion; that traffic congestion has a strong effect on driver's visual characteristics has been proved via adequate data analysis; the risky lane changing behaviors in traffic congestion have been specifically analyzed.

The comparative analysis of the dynamic visual characteristics (fixation points' distribution, fixation duration, average saccade speed, average saccade acceleration, blink duration, and blink rate) is done via the data being obtained during rush hour. And, it is discovered that the eye movement characteristics are different between the general and congested state.

The characteristics of the risky lane changing behaviors are obtained through video monitoring, and the risky lane changing behaviors on the pressure of congestion have been divided into three types, and the proportion of the three types

of risky lane changing behaviors is analyzed by statistical tools. Further, the linear relationship between the number of the traffic conflicts and risky lane changing is established to measure the security features for the risky lane changing in the pressure of traffic congestion.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The study was supported by the National Natural Science Foundation of China (51078113 and 51178149).

## References

- [1] J. D. Sun, Q. Liu, and Z.-R. Peng, "Research and analysis on causality and spatial-temporal evolution of urban traffic congestions—a case study on Shenzhen of China," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 5, pp. 86–93, 2011.
- [2] R. Arnott, "A bathtub model of downtown traffic congestion," *Journal of Urban Economics*, vol. 76, pp. 110–121, 2013.
- [3] T. Tsekeris and N. Geroliminis, "City size, network structure and traffic congestion," *Journal of Urban Economics*, vol. 76, pp. 1–14, 2013.
- [4] S. Yang, "On feature selection for traffic congestion prediction," *Transportation Research Part C*, vol. 26, pp. 160–169, 2013.

- [5] B. L. Smith, B. M. Williams, and R. K. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C*, vol. 10, no. 4, pp. 303–321, 2002.
- [6] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.
- [7] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *Journal of Intelligent Transportation Systems*, vol. 13, no. 2, pp. 53–72, 2009.
- [8] D. A. Hennessy and D. L. Wiesenthal, "The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours," *Ergonomics*, vol. 40, no. 3, pp. 348–361, 1997.
- [9] D. A. Hennessy and D. L. Wiesenthal, "Traffic congestion, driver Stress, and driver aggression," *Aggressive Behavior*, vol. 25, no. 6, pp. 409–423, 1999.
- [10] B.-H. Liu, L.-S. Sun, and J. Rong, "Driver's visual cognition behaviors of traffic signs based on eye movement parameters," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 4, pp. 22–27, 2011.
- [11] T. C. Lansdown, "Individual differences during driver secondary task performance: verbal protocol and visual allocation findings," *Accident Analysis and Prevention*, vol. 34, no. 5, pp. 655–662, 2002.
- [12] G. Underwood, D. Crundall, and P. Chapman, "Selective searching while driving: the role of experience in hazard detection and general surveillance," *Ergonomics*, vol. 45, no. 1, pp. 1–12, 2002.
- [13] S. Benedetto, M. Pedrotti, L. Minin, T. Baccino, A. Re, and R. Montanari, "Driver workload and eye blink duration," *Transportation Research Part F*, vol. 14, no. 3, pp. 199–208, 2011.
- [14] J. L. Deffenbacher, R. S. Lynch, E. R. Oetting, and D. A. Yingling, "Driving anger: correlates and a test of state-trait theory," *Personality and Individual Differences*, vol. 31, no. 8, pp. 1321–1331, 2001.
- [15] J. L. Deffenbacher, R. S. Lynch, E. R. Oetting, and R. C. Swaim, "The driving anger expression inventory: a measure of how people express their anger on the road," *Behaviour Research and Therapy*, vol. 40, no. 6, pp. 717–737, 2002.
- [16] M. J. M. Sullman, "Anger amongst New Zealand drivers," *Transportation Research Part F*, vol. 9, no. 3, pp. 173–184, 2006.
- [17] D. J. Rapport and A. M. Friend, *Towards a Comprehensive Framework for Environmental Statistics: A Stress-Response Approach*, Statistics Canada, Ottawa, Canada, 1979.
- [18] D. J. Rapport and A. Singh, "An EcoHealth-based framework for State of Environment Reporting," *Ecological Indicators*, vol. 6, no. 2, pp. 409–428, 2006.
- [19] D. Zhou, Z. Lin, L. Liu, and D. Zimmermann, "Assessing secondary soil salinization risk based on the PSR sustainability framework," *Journal of Environmental Management*, vol. 128, pp. 642–654, 2013.
- [20] K. Lundberg, B. Balfors, and L. Folkesson, "Framework for environmental performance measurement in a Swedish public sector organization," *Journal of Cleaner Production*, vol. 17, no. 11, pp. 1017–1024, 2009.
- [21] R. W. Novaco, D. Stokols, and L. Milanese, "Objective and subjective dimensions of travel impedance as determinants of commuting stress," *American Journal of Community Psychology*, vol. 18, no. 2, pp. 231–257, 1990.
- [22] E. Gulian, G. Matthews, A. I. Glendon, D. R. Davies, and L. M. Debney, "Dimensions of driver stress," *Ergonomics*, vol. 32, no. 6, pp. 585–602, 1989.

## Research Article

# The Effects of the Tractor and Semitrailer Routing Problem on Mitigation of Carbon Dioxide Emissions

Hongqi Li,<sup>1</sup> Yanran Li,<sup>1</sup> Yue Lu,<sup>1</sup> Qiang Song,<sup>2</sup> and Jun Zhang<sup>3</sup>

<sup>1</sup> School of Transportation Science and Engineering, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing 100191, China

<sup>2</sup> School of Humanities and Social Sciences, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing 100191, China

<sup>3</sup> School of Economics and Management, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing 100191, China

Correspondence should be addressed to Hongqi Li; [lihongqi@buaa.edu.cn](mailto:lihongqi@buaa.edu.cn)

Received 19 August 2013; Revised 30 September 2013; Accepted 24 October 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Hongqi Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The incorporation of CO<sub>2</sub> emissions minimization in the vehicle routing problem (VRP) is of critical importance to enterprise practice. Focusing on the tractor and semitrailer routing problem with full truckloads between any two terminals of the network, this paper proposes a mathematical programming model with the objective of minimizing CO<sub>2</sub> emissions per ton-kilometer. A simulated annealing (SA) algorithm is given to solve practical-scale problems. To evaluate the performance of the proposed algorithm, a lower bound is developed. Computational experiments on various problems generated randomly and a realistic instance are conducted. The results show that the proposed methods are effective and the algorithm can provide reasonable solutions within an acceptable computational time.

## 1. Introduction and Problem Description

Along with the growth of the demand for goods, transportation volume is increasing rapidly. For door-to-door transportation, the most widely used mode is road transportation. However, road transportation has some negative impact on the environment because of the land use, energy resource consumption, and so forth. Road transportation accounts for almost 80% of total energy demands from transportation [1]. Fossil fuels are the main energy sources of transportation and CO<sub>2</sub> is emitted during the combustion of fossil fuels. As a dominant mode of freight movement, road transportation accounts for the largest share of the freight-related emissions [2]. The percentages of road freight transportation CO<sub>2</sub> emissions compared to the entire transportation sector from 1985 to 2007 in China were between 29% and 34% [3]. More and more tons of CO<sub>2</sub> are released into the environment annually and the road freight transportation CO<sub>2</sub> emissions are likely to keep growth. Recent studies on freight transportation have focused not only on cost minimization or profit maximization for a freight company but also on carbon

reduction to enhance the corporate social responsibility for the company [4].

The fuel consumption is the most expensive variable cost of the transportation process to road freight transportation enterprises [5]. The efficient use of trucks and road networks becomes more and more important. Compared with costly network infrastructure modifications, optimized routing strategies have been proven to be more efficient in enhancing network capacity [6]. Optimized vehicle routing can reduce the number of trucks and utilize the network better by reducing vehicle movements. The optimization problem has been extensively studied in the literature, known as the vehicle routing problem (VRP). The VRP and its various extensions have long been one of the most studied combinatorial optimization problems due to the problem's complexity and extensive applications [7, 8].

The incorporation of the minimization of energy and CO<sub>2</sub> emissions in the VRP is a relatively recent topic addressed in the research work. VRP-related research that aims to minimize total fuel consumption is rather rare [9]. Kuo [9] proposed a model for calculating total fuel consumption for

the time-dependent VRP. A SA algorithm was proposed. The experimental results showed that there may be a trade-off between fuel consumption, transportation time, and transportation distances. Pradenas et al. [10] presented a mathematical model for vehicle routing with time windows and backhauling. A Scatter Search (SS) metaheuristic that minimized the emission of greenhouse gases for a homogeneous vehicle fleet was designed and implemented. The experimental results showed that the environmental impacts of the vehicle routing are related to the vehicle use rate, the loads that are transported among customers, and the distances between customers.

In general, various types of vehicles are used by enterprises. Through the constitution mode of the autonomous part and the nonautonomous part [11], vehicles can be classified into two main types: trucks and combination vehicles. A truck, that is, a single-unit truck, has its fixed autonomous and nonautonomous parts and the two parts cannot be separated. Within the class of combination vehicles are truck tractor-semitrailer combinations and trucks or truck tractors with semitrailers in combination with full trailers. In some extensions of the VRP, the combination vehicles, especially the use of trailers (a commonly neglected feature in the VRP), are considered. We study a variant of the combination-vehicle routing problem where tractor-semitrailer combinations are utilized. We call this variant of the VRP as the tractor and semitrailer routing problem (TSRP). As concern over global warming has grown, reducing CO<sub>2</sub> emissions is becoming important for transportation. Minimizing fuel consumption will be increasingly influential in the VRP. Differing from other literature, we aim to minimize fuel consumption of the TSRP through considering the vehicle type and its use rate.

The TSRP in the paper has the following characteristics.

- (1) A tractor of the tractor with semitrailer combination can pull independent semitrailers. The tractor cannot load goods and it is only used for pulling semitrailers. The time to attach/detach a semitrailer to a tractor at a location is usually considerably less than the time to load/unload all cargoes in a semitrailer, so the tractor in the TSRP has a high use rate.
- (2) The transportation service provider serves the customer orders from a number of depots. Shipments occur between depots and the pickup/delivery locations of an order, and between depots. Either full truckload (TL) or less than truckload (LTL) shipments may exist. Trucks considered in the VRP are fit for LTL shipments. In the TSRP, semitrailers are fit for TL.
- (3) The tractor and semitrailer transportation is promoted to be more energy-efficient than single-unit trucks transportation [12]. As mentioned by Ierland et al. [13], the CO<sub>2</sub> emission factor, which can be defined as CO<sub>2</sub> emissions per ton-kilometer (unit: g CO<sub>2</sub>/t-km), is a typical index to describe the CO<sub>2</sub> emission effects of the road freight transportation.
- (4) One of the important applications of the TSRP is the concept of multilevel freight distribution systems

(e.g., city logistics and multimodal freight transportation systems), in which freight arrives at a central depot and is transported further to satellite facilities by larger vehicles, and the freight is then brought to the final customers by smaller vehicles. The problem of how to efficiently route vehicles operating at both levels is known in the literature as the Two-Echelon Vehicle Routing Problem (2E-VRP) [14], the generalized vehicle routing problem (GVRP) [15], or the single-sourcing two-echelon capacitated location-routing problem (2E-CLRP) [16]. The tractor and semitrailer combination has high average loads and a high use rate of tractors, which makes it feasible to be used in the level where shipments are large or TL.

We propose the TSRP on a loaded-semitrailer flow network. There are two types of terminals on the network: one central depot and a number of satellite facilities. At the beginning, all tractors locate in the central depot, while the satellite facilities have loaded-semitrailers waiting for sending. All tractors or vehicles (a vehicle is one tractor pulling one semitrailer) originate and terminate at the central depot. A homogeneous fleet composed of tractors and semitrailers serves the flow demand among terminals. A tractor can pull one loaded-semitrailer and can also run alone. The loaded-semitrailers are assumed TL. The objective of the TSRP is to determine the number of tractors and the route of each tractor so as to minimize CO<sub>2</sub> emissions per ton-kilometer.

There are some features that distinguish the TSRP from existing research on routing problems. Firstly, we extend the application background of tractor and semitrailer combinations to the TL transportation course of multilevel freight distribution systems. The application of the TSRP is different from tractor and semitrailer combination applications in the literature, which is called the roll-on roll-off vehicle routing problem (RRVRP). Secondly, the TSRP takes a composite index, that is, CO<sub>2</sub> emissions per ton-kilometer, as the objective. It is different from most of the VRPs which take single index (e.g., total distance, total cost, etc.) as the objective. In practice, statistical data on CO<sub>2</sub> emissions per ton-kilometer can be used to calibrate the experimental results of the TSRP. Thirdly, the nodes of the TSRP may send more than one loaded-semitrailer and the nodes and linked arcs must be visited more than once. A node may appear more than once in the same route of the solution.

Our interest in the TSRP arises from real-life regional-level truck operations in China and the TL transporting of multilevel freight distribution systems. Our aim is to develop a solving method for the TSRP and to demonstrate the effect of CO<sub>2</sub> emission mitigation. The paper is organized as follows. The next section introduces the relevant literature. A mathematical model and a lower bound for the TSRP are developed in Section 3. Section 4 proposes the heuristic algorithm for solving the problem. Computational experiments are described in Section 5. Finally, conclusions and future work are given in Section 6.

## 2. Literature Review

In this section we turn to an overview of contributions to the combination-vehicle routing problem the TSRP addresses.

Research on the VRP to date has considered especially trucks, truck and full trailer combinations. The truck and trailer routing problem (TTRP) has been brought forward for decades. In the TTRP, a heterogeneous fleet composed of trucks and truck and full trailer combinations serves a set of customers. Each customer has a certain demand, and the capacities of the trucks and trailers are determinate. Some customers must be served only by a truck, while other customers can be served either by a truck or by a combination vehicle. The objective of the TTRP is to find a set of routes with minimum total distance or cost so that each customer is visited in a route. Chao [17] distinguished three different types of routes in a TTRP solution: a pure truck route (PTR) where a truck serves all customers without using a trailer, a pure vehicle route (PVR) where all customers are served by a truck with a coupled trailer, and a complete vehicle route (CVR) where at least once the trailer is uncoupled from the truck at a vehicle customer and the truck continues serving a subset of customers on a truck subroute.

Semet and Taillard [18] and Gerdessen [19] studied the TTRP in the 1990s. Semet and Taillard [18] and Caramia and Guerriero [20] gave some real-world TTRP applications. Gerdessen [19] extended the VRP to the vehicle routing problem with trailers and investigated the optimal deployment of a fleet of truck-trailer combinations. Scheuerer [21] proposed construction heuristics along with a Tabu search algorithm for the TTRP. Tan et al. [22] proposed a hybrid multiobjective evolutionary algorithm to solve the TTRP. Lin et al. [23] proposed a simulated annealing (SA) heuristic for the TTRP. Villegas et al. [24] solved the TTRP by using a hybrid metaheuristic. On the extensions of the TTRP, Villegas et al. [25] proposed two metaheuristics to solve the single truck and trailer routing problem with satellite depots (STTRPSD). Considering the number of available trucks and trailers being limited in the TTRP, Lin et al. [26] relaxed the fleet size constraint and developed a SA heuristic for solving the relaxed truck and trailer routing problem (RTTRP). Lin et al. [27] proposed a SA heuristic for solving the truck and trailer routing problem with time windows (TTRPTW). Recently, Derigs et al. [28] combined local search and large neighborhood search metaheuristic to solve the TTRP with/without load transfer and the TTRP with/without time windows. Computational tests on benchmark instances showed that the approach was at least competitive to state-of-the-art approaches for the TTRP without time windows.

In the TTRP, each trailer can be pulled by a unique associated truck, and only this truck is permitted to transfer the load into the trailer. The amount of trucks is generally more than that of trailers. Drexel [11] described the VRP with trailers and transshipments (VRPTT) in which there is no fixed truck-trailer assignment. The TTRP is evidently a special case of the VRPTT. Besides, Pureza et al. [29] addressed the VRP with time windows and multiple delivery men (VRPTWMD) that allows a number of delivery men to be assigned to each route. Two solution approaches based on

Tabu search and ant colony optimization were proposed. The impact of the use of extra delivery men in route planning was assessed by means of computational experiments. If regarding delivery men as trucks, the VRPTWMD becomes the TTRP when there is only one deliveryman.

There are several variants of the VRP which consider tractor and semitrailer combinations. These variants include the RRVRP and others. In the literature, the RRVRP arises when tractors move large trailers between locations generating a high volume of waste like construction sites and disposal facilities. In the basic RRVRP, there is a single depot where all tractors are located at the beginning. There is a single disposal facility where full trailers are dumped and empty trailers can be put on or pulled from inventory. At the end of the day, all tractors return to the depot while trailers may remain at customer locations or the disposal facility. The problem is to assign trips to tractors and to find routes for the tractors that do not exceed a given maximal duration and that minimize the nonproductive deadhead time of tractors between trips as well as the number of tractors used.

Bodin et al. [30] studied the RRVRP with a depot and a disposal site and classified customer demands into four trip types. They defined the RRVRP as a combination of an asymmetric vehicle routing problem with a bin packing problem. Heuristic methods were proposed to solve some benchmark problems on the RRVRP. Derigs et al. [31] solved the RRVRP by combining local search and large neighborhood search controlled by two relatively simple and parameter-free/-poor metaheuristic control procedures. Wy et al. [32] introduced the RRVRP with time windows (RRVRPTW). The objective of the RRVRPTW is to minimize the number of required tractors and their total route time. A LNS based iterative heuristic approach consisting of a construction algorithm and several improvement algorithms was proposed. Baldacci et al. [33] modeled the multiple disposal facilities and multiple inventory locations RRVRP (M-RRVRP) as a time constrained vehicle routing problem on a multigraph.

There are other variants concerning the tractor and semitrailer combination routing problem. Hall and Sabnani [34] studied routes that consisted of two or more segments and two or more stops in the route for a tractor. Control rules based on predicted route productivity were developed to determine when to release a tractor. Francis et al. [35] solved the multiresource routing problem (MRRP) with flexible tasks. Two resources (tractors and trailers) performed tasks to transport loaded and empty equipment. Cheng et al. [36] proposed a model for a steel plant to find the tractor and semitrailer running routes for the purpose of minimizing transport distance. Derigs et al. [37] presented two approaches to solve the vehicle routing problem with multiple uses of tractors and trailers. Li et al. [38] studied the tractor and semitrailer routing problem on a unit-flow network, and a heuristic algorithm was used to decide the number of tractors and the route of each tractor.

In the literature, combination-vehicle routing problem is becoming hot in recent years. The background of the TTRP applications is similar to that of the VRP, that is, city logistics or other delivery process. In fact, an intercity line-haul tier is

necessary to perfect city logistics or other delivery systems. Bulk transportation of large volumes of freight between cities allows economies of scale to be achieved by using large-capacity vehicles. Although the use of tractor and semitrailer combinations is considered by the RRVRP in the literature, the RRVRP applications are mainly limited in waste collection business. Because of high use rate of tractors, tractor and semitrailer combinations are promoted to be more energy-efficient than single-unit trucks. There is another variant of the RRVRP when the application background becomes intercity line-haul transportation. Among all the work we have reviewed in literature, little work has been done on the TSRP we have described earlier. The problem considered in the present study involves incorporating CO<sub>2</sub> emissions minimization in the TSRP. In addition, the transportation enterprise can reject customer order either because serving the order is impossible or because the cost of serving the order is too high.

### 3. Model Formulation

The underlying assumptions of the TSRP model are (i) all loaded-semitrailer flow demands are known in advance. The problem is static. Empty semitrailer exchanges are ignored. (ii) Loaded-semitrailer flow demands can originate between any two terminals. At the central depot, the number of incoming semitrailers is equal to the number of outgoing semitrailers; that is, the central depot has balanced flows. (iii) Some loaded-semitrailer flow demands may be rejected. The transportation service level is based on the percentage of flow demands that are satisfied. (iv) A route must not exceed a given time-span. In order to balance the route lengths, a route has to exceed a given minimization-time. If a satellite facility is already in a route, it cannot be reinserted in the same route. (v) All tractors are assigned to the central depot where they must return to after each route. Routes must start from and end at the central depot. Each tractor leaves from and returns to the depot exactly once. (vi) The TSRP deals with TL and does not consider cross-docking options.

The TSRP can be formulated as follows.

Let  $G = (V, A)$  be a directed graph where  $V = \{0, 1, 2, \dots, n\}$  is the vertex set and  $A = \{(i, j) \mid i, j \in V, i \neq j\}$ , the arc set. Vertex 0 ( $v_0$ ) is the central depot and the other vertices ( $v_i$ ) in  $V$  (i.e.,  $V \setminus \{0\}$ ) correspond to satellite facilities. Loaded-semitrailer flows between any two terminals are  $R$ :

$$R = \begin{bmatrix} 0 & r_{01} & r_{02} & \cdots & r_{0n} \\ r_{10} & 0 & r_{12} & \cdots & r_{1n} \\ r_{20} & r_{21} & 0 & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n0} & r_{n1} & \cdots & r_{n,n-1} & 0 \end{bmatrix}, \quad (1)$$

where  $r_{ij}$  denotes that there are  $r_{ij}$  loaded-semitrailers needed to be transported from  $v_i$  to  $v_j$ .

$X_{ijk}^t$  and  $X_{ijk}^l$  are the decision variables. If the  $k$ th tractor runs from  $v_i$  to  $v_j$ ,  $X_{ijk}^t = 1$ ; otherwise,  $X_{ijk}^t = 0$ . If the  $k$ th loaded combination vehicle (i.e., the  $k$ th tractor pulling one

loaded-semitrailer) runs from  $v_i$  to  $v_j$ ,  $X_{ijk}^l = 1$ ; otherwise,  $X_{ijk}^l = 0$ .

$C_{ijk}^t$  is the fuel consumption of the  $k$ th tractor running from  $v_i$  to  $v_j$ .  $C_{ijk}^l$  is the fuel consumption of the  $k$ th loaded combination vehicle running from  $v_i$  to  $v_j$ .  $T_{ijk}^t$  is the running time of the  $k$ th tractor running from  $v_i$  to  $v_j$  and  $vel^t$  the average velocity.  $T_{ijk}^l$  is the running time of the  $k$ th loaded combination vehicle running from  $v_i$  to  $v_j$  and  $vel^l$  the average velocity. The tractor's on-duty hours per day is  $T_0$ .

The objective function is

$$\text{Min } \gamma \cdot \frac{(\sum_i \sum_j \sum_k C_{ijk}^t X_{ijk}^t + \sum_i \sum_j \sum_k C_{ijk}^l X_{ijk}^l)}{(W \cdot \sum_k \sum_i \sum_j X_{ijk}^l \cdot T_{ijk}^l \cdot vel^l)} \quad (2)$$

The constraints are

$$\sum_k \sum_i \sum_j X_{ijk}^l \geq \eta \cdot \sum_i \sum_j r_{ij}, \quad (3)$$

$$\sum_k X_{ijk}^l \leq r_{ij}, \quad (4)$$

$$\sum_i X_{0ik}^t + \sum_i X_{0ik}^l = 1, \quad (5)$$

$$\sum_i X_{i0k}^t + \sum_i X_{i0k}^l = 1, \quad (6)$$

$$\left( \sum_i \sum_k X_{ij'k}^t + \sum_i \sum_k X_{ij'k}^l \right) - \left( \sum_i \sum_k X_{j'ik}^t + \sum_i \sum_k X_{j'ik}^l \right) = 0, \quad (j' = 1, 2, \dots, n), \quad (7)$$

$$X_{ijk}^t + X_{ijk}^l \leq 1, \quad (8)$$

$$\sum_i \sum_k X_{i0k}^l - \sum_i \sum_k X_{0ik}^l = 0, \quad (9)$$

$$X_{ijk}^t + X_{j'hk}^t \leq 1, \quad (10)$$

$$\sum_i \sum_j T_{ijk}^l \cdot X_{ijk}^l > \sum_i \sum_j T_{ijk}^t \cdot X_{ijk}^t, \quad (11)$$

$$t_H + T_{st} + \sum_i \sum_j (T_{ijk}^t \cdot X_{ijk}^t + T_{ijk}^l \cdot X_{ijk}^l) \geq \rho_1 \cdot T_0, \quad (12)$$

$$t_H + T_{st} + \sum_i \sum_j (T_{ijk}^t \cdot X_{ijk}^t + T_{ijk}^l \cdot X_{ijk}^l) \leq \rho_2 \cdot T_0. \quad (13)$$

$X_{ijk}^t \in \{0, 1\}$ ;  $X_{ijk}^l \in \{0, 1\}$ ;  $i, j, h = 0, 1, 2, \dots, n$ ;  $i \neq j$ ;  $j \neq h$ ;  $k$  is an integer and the maximum  $k$  is the total of tractors used.

The objective function is the CO<sub>2</sub> emissions per ton-kilometer of the TSRP, where  $\gamma$  is the emission coefficient (a constant) and  $W$  the freight weight on a loaded-semitrailer.

Constraints (3) and (4) guarantee that the total freight flow demand is satisfied by a certain percentage ( $\eta$ ,  $0 < \eta \leq 1$ ), while the demand of any terminal is respected. Constraints (5) and (6) guarantee that any tractor starts from the central depot and terminates at the central depot. Constraints (7) guarantee that satellite facilities cannot reserve any tractor. Constraints (8) are the restrictions of the tractor (i.e., running-alone tractor and tractor attached in combination vehicle) passing by a certain arc. Constraints (9) guarantee the flow balance of the central depot. Constraints (10) forbid the illegal segments of routes. Constraints (11) guarantee the load-running rate of tractors.

Constraints (12) and (13) are the restrictions of balancing the route lengths, where  $\rho_1$  ( $0 < \rho_1 < 1$ ) and  $\rho_2$  ( $1 \leq \rho_2 \leq \tau$ ,  $\tau$  is a limited number) are the lower and upper limits of the utilization ratio of the tractor's on-duty hours, respectively.  $t_H$  is the tractor's residence time at the central depot.  $T_{st}$  is the total of tractors' attach/detach time at satellite facilities.  $T_{st}$  is affected by the number of satellite facilities included in a route. When a route  $k'$  includes  $f_{k'}$  satellite facilities and the tractor's attach/detach time at a satellite facility is  $t_{sf}$ ,  $T_{st} = f_{k'} \cdot t_{sf}$ , where  $f_{k'} = \sum_i \sum_j (X_{ijk}^t + X_{ijk'}^l) - 1$ .

Although the decision variables of the formulation are binary, the route is decided. Constraints (5)~(10) suggest that the route is made up of sequential arcs which are passed by the tractor or the loaded combination vehicle. We denote the terminal sequence of a route by  $H - s_1 - \dots - s_f - H$ , where  $H$  is the central depot and  $s_i$  satellite facility.

Generally, efficient exact algorithms to solve the model we presented here for realistic problem sizes do not exist. Thus, such model can only be solved by heuristics to attain suboptimal solutions of a priori unknown quality. In such cases, it is useful to find lower bounds to get an estimate for the quality of the solution found by the heuristics. If the number of variables and constraints in a derivative model is significantly less than those in the original model, the derivative model is expected to be solved in a much shorter time. This derivative model is solved to find lower bounds on the objective values of the solutions to the original model [39]. If we relax the route length constraints (12) and (13) in the model above, the objective function and constraints (3)~(11) can be reformulated as another model (denoted as LBM) that can be solved on small-scale instances. The solution of LBM may also be the optimal solution of an instance if it satisfies the relaxed time-span constraints. We use this lower bound to compare the performance of the heuristics presented in Section 4.

In addition, the objective function of the TSRP model is substantially affected by two parts: one ( $(\sum_i \sum_j \sum_k C_{ijk}^t X_{ijk}^t) / (\sum_k \sum_i \sum_j X_{ijk}^l \cdot T_{ijk}^l \cdot \text{vel}^l)$ ) and the other ( $(\sum_i \sum_j \sum_k C_{ijk}^l X_{ijk}^l) / (\sum_k \sum_i \sum_j X_{ijk}^l \cdot T_{ijk}^l \cdot \text{vel}^l)$ ). When constraints (3) and (4) are equations and the satisfied demand percentage is 100%, the objective function is affected by the former part. If any segment on routes has loaded-semitrailer transported, that is,  $X_{ijk}^t = 0$  for all  $i, j$ , and  $k$ , the objective function reaches the minimum. The minimum is suggested as a benchmark for LBM.

## 4. Heuristic Algorithm

Three types of algorithms are used to solve the VRP [40]. The first type consists of exact algorithms that are time-consuming. The second type consists of classical heuristics such as greedy, local search, and relaxation based. The third type consists of heuristics that are based on some metaheuristic rules. Such metaheuristics or framework for building heuristics are SA, Tabu search, Genetic algorithms, Variable neighborhood search, and so forth. The high computational cost of exact methods and their poor performance in large problems have involved that the current research concentrates on stochastic algorithms that are capable of producing feasible but not necessarily optimal solutions in limited time [5]. The SA algorithm is one of the commonly used metaheuristics, which has been successfully applied to solve several types of VRP. Motivated by the success of the SA for the TTRP (e.g., [26, 27]), we have therefore opted a heuristic algorithm based on the SA to solve the TSRP.

The SA uses a stochastic approach to search for and move to neighborhood solutions. If a better neighborhood solution is identified in the search starting from the current solution, the move will be accepted and the current solution will be replaced by the better neighborhood solution. The search for a better neighborhood solution then continues. Besides, the SA will accept the moving to a worse neighborhood solution with a certain probability to escape from a local optimum. The accepted probability is based on two parameters, the temperature which gradually reduces and the objective function difference between the two solutions. At the beginning of the search, the accepted probability of the move is higher. When nearing the end of the search process, the accepted probability of the move is smaller. Generally, the initial temperature, the cooling function, and the final temperature will affect the results of the SA.

*4.1. The Neighborhood and Initial Solution.* Braysy and Gendreau [41] stated that local search plays a very important role in the design of metaheuristics for the VRP. A local search operator iteratively improves a solution by exploring its neighborhood. The TSRP model suggests that the route is made up of sequential arcs. Any ones of the constraints of (5) to (10) guarantee some requirements for segments of a route, while the route length constraints (12) and (13) guarantee an entire requirement of a route. We regard constraints (12) and (13) as the most important factors to decide a solution's neighborhood.

The tractor's on-duty hours per day ( $T_0$ ) depend on the on-duty hours of the driver team. A driver's on-duty hours (denoted by  $T_p$ ) consist of driving hours, attach/detach time ( $t_{sf}$ ) and residence time ( $t_H$ ) at terminals. If the number of drivers allocated to each tractor is  $h$ ,  $T_0 = h \cdot T_p$ . When the transporting distance between terminals  $v_i$  and  $v_j$  is  $d_{ij}$  and  $\text{vel}$  the average velocity, the running time between  $v_i$  and  $v_j$  is ( $d_{ij}/\text{vel}$ ). Constraints (12) and (13) are rewritten as

$$\rho_1 \cdot h \cdot T_p \leq \sum_i \sum_j \left( \frac{d_{ij}}{\text{vel}} \right) + f_{k'} \cdot t_{sf} + t_H \leq \rho_2 \cdot h \cdot T_p. \quad (14)$$

The vehicle-kilometer (v-km) is the most important index for enterprises to evaluate the performance of vehicles and drivers. When there are  $f_{k'}$  satellite facilities in a route, the more the  $f_{k'}$ , the slower the average velocity because the tractor/vehicle has to frequently enter into and depart off terminals where a slow velocity is allowed. The average velocity (denoted by  $\text{vel}(f_{k'})$ ) is regarded as a decreasing function. If the utilization ratio of the tractor's on-duty hours is 100%, the running time of the route is  $(h \cdot T_p - f_{k'} \cdot t_{sf} - t_H)$ . Then, the v-km of vehicles/drivers is  $VK = \text{vel}(f_{k'}) \cdot (h \cdot T_p - f_{k'} \cdot t_{sf} - t_H)$ . The v-km will decrease along with increasing  $f_{k'}$ . Therefore, the routes include finite inserted satellite facilities. We enumerate the number of satellite facilities in a route to search entirely to find all routes that satisfy constraints (14). When the routes include at least 1 and at most  $f_{k'}$  satellite facilities, there are  $O(n^{f_{k'}})$  potential routes for selection, which also make up the neighborhood.

Unlike the VRP or the TTRP where every removal is accompanied by an insertion, one may decide removal and not continue an insertion in our problem. This partial solution will still be feasible when the demand satisfied percentage keeps larger than the minimum ( $\eta$ ). By referring to the traditional destroy and repair framework, we take a whole route as the operator unit. There are three types of operators. (i) A route is removed from the current solution by a destroy operator. (ii) A route is removed from the current solution by a destroy operator and another route is reinserted by a repair operator. The removed route is recorded by the neighborhood and is still a candidate route of the repair operator. (iii) A route may clone itself several times. The times are decided by the maximum demand of satellite facilities included by the route. The clone operator is a special type of repair operator. The realistic-size instance results in Section 5 show the effect of the clone operator.

To the generation of the initial solution, our computational tests show a similar conclusion as that of Coelho et al. [42]. The initial solution does not have significant impacts on the overall solution cost or the running time. We therefore generate randomly the initial solution.

**4.2. The SA Heuristic.** The number of tractors (or the number of routes in the solution) is unknown, but it is an essential parameter at the beginning. It is likely that a tractor pulls more than one independent semitrailer on the route. Denoting the average number of transported loaded-semitrailers on a route as  $\xi$ ,  $(\sum_i \sum_j r_{ij})/\xi$  is an important benchmark of the number of tractors. The number of tractors should be limited integers. Our experimental results show that the CO<sub>2</sub> emissions per ton-kilometer can be decreased by adding or subtracting up to  $(\sum_i \sum_j r_{ij})/(3 \cdot \xi)$  extra tractors. It is suggested that the SA heuristic for the TSRP does not take the number of tractors as one of the objective values. Meanwhile, the computational tests show that it is feasible for  $\xi = 4$ . The SA procedure is started by selecting randomly  $N_0 = \lfloor (\sum_i \sum_j r_{ij})/4 \rfloor$  routes as the initial solution where  $\lfloor \bullet \rfloor$  denotes the largest integer which is smaller than or equal to the enclosed number. In each iteration, the objective function

which is used to evaluate the sequential solutions is the percentage of satisfied freight demand. When the percentage of satisfied freight demand reaches  $\eta$ , the SA procedure finds a nondominated solution for the number of tractors. If the percentage of satisfied freight demand is less than  $\eta$ , the number of routes in the former solution ( $N_i$ ) is enlarged by  $(1 + \delta)$  times (where  $0 < \delta < 0.05$ ) and the SA procedure started with  $\lfloor N_i \cdot (1 + \delta) \rfloor$  routes. Finally, the number of tractors is decided.

The SA procedure is run in two phases. In the first phase, a nondominated solution with the minimum number of tractors is obtained. In the second phase, successively CO<sub>2</sub> emissions per ton-kilometer are minimized for the current number and for the number of tractors which is increased by one every time until a predefined stopping criterion is met (e.g., the CO<sub>2</sub> emissions per ton-kilometer stop decrease).

The initial temperature, the final temperature, the Boltzmann constant used in the probability function, the maximum number of iterations, and the cooling mechanism adjusted by the number of iterations are made certain firstly. The probability function used for the SA procedure deciding the number of tractors is different from that for the SA procedure deciding the final solution of the TSRP. The former mainly considers the percentage of satisfied freight demand, while the latter considers synchronously the percentage of satisfied freight flow demand and the CO<sub>2</sub> emissions per ton-kilometer. As proposed by Kuo [9], our SA model also involves the temperature being cooled each iteration, which is different from general SA where a certain number of iterations take place between each cooling. We use an initial temperature of 100, a final temperature of 1 and the number of iterations in the SA equals to 100000 (for the small-scale random instances of the next section) or 4500000 (for the realistic instance of the next section). The cooling mechanism is adjusted based on the number of iterations. The termination condition is to stop the algorithm when the number of iterations reaches its maximum and the temperature becomes 1.

## 5. Computational Study

Since we are not aware of any prior test instance for the TSRP minimizing CO<sub>2</sub> emissions per ton-kilometer, the proposed model and algorithm were tested on a range of small-scale instances generated randomly and a realistic instance. Our computational experiments were carried out in two parts. First, the small-scale instances were used to show the effectivity of our model and the heuristic. We have also calibrated the solution methods on small-scale instances as explained in Section 5.1. Second, we have run the heuristic on a realistic instance with varying central depot location as explained in Section 5.2. The emphasis lies on summing some calibrating methods to seek a closely parameter-free metaheuristic by small-scale instances generated randomly. Referring to calibrating methods, the SA heuristic algorithm can generate high-quality solutions in relatively little time with some parameter tuning.

**5.1. Small-Scale Instances.** The proposed TSRP model and the SA heuristic are assessed on a number of small-scale test problems on  $5 \times 5$  grid graphs. Several parameters are considered while the problems are created (1) the number of satellite facility nodes ( $C_n$ ); (2) the number of loaded-semitrailers of each satellite facility; (3) the location of the central depot; (4) the location of the satellite facilities; (5) the parameters of tractors and semitrailers (e.g., fuel consumption, cargo weight of a loaded-semitrailer, velocity, etc.); (6) the distance between any two nodes; (7) the distance span of a route.

The small-scale instances are divided into four sets according to the number of satellite facilities. The small-scale instances are created by a random fashion as follows. The “RANDOM” function in Matlab is used. By RANDOM (“norm”,1,1,5,5), random arrays are generated. We select the minimum position of a random array as the central depot and other  $C_n$  negative positions of the array satellite facilities. The distance between any two terminals is calculated by the gaps of rows and columns. The “RANDOM” function is used ten times and each set includes ten instances. The “RANDOM” function is also used to determine the number of loaded-semitrailers of terminals.

The tractor-semitrailer combination, which can load maximally 30 tons and satisfies the fuel-efficiency requirements of “Regulation of Supervising Vehicle Fuel Consumption” (number 11/2009 Decree of Ministry of Transport of the People’s Republic of China (MOTPRC)), is used to transport goods. The type code of the selected combination is “CQ4254HTVG324V” or “ND4251B32J7.” The fuel consumption is 18 liters diesel per 100 kilometers for a tractor running alone and 32 liters diesel per 100 kilometers for a combination run. Suppose that: the gap between adjacent rows or columns of the random array is 50 km; the loading factor of loaded-semitrailer is 60%; the velocity is 50 km/h; the distance span is around 650 km; and the percentage of satisfied freight demand is not less than 85%.

The test problems are detailed in Table 1. Columns 1–9 indicate the test problem, the rectangle region, the total number of loaded-semitrailers, the average loaded-semitrailer number of all satellite facilities, the variance of loaded-semitrailer number of all satellite facilities, the minimum distance between central depot and satellite facility, the maximum distance between central depot and satellite facility, the average distance between any two terminals and the variance of distances between any two terminals respectively.

To obtain lower bounds, the integer programming model (LBM) presented in Section 3 has been implemented and solved using LINGO11. The solving course of LBM took the number of tractors as a precondition; that is,  $k = \lfloor (\sum_i \sum_j r_{ij})/4 \rfloor$ . When computing the lower bound LBM for an instance, we may find a better value of  $k$  during the computational experiments. In course of finding feasible solution for LBM, the number of tractors is adjusted manually. Besides, the proposed SA algorithm has been coded in MATLAB R2010b and run on a computer with an AMD Athlon(tm) X2 Dual-Core QL-65 running at 2.10 GHz under Windows

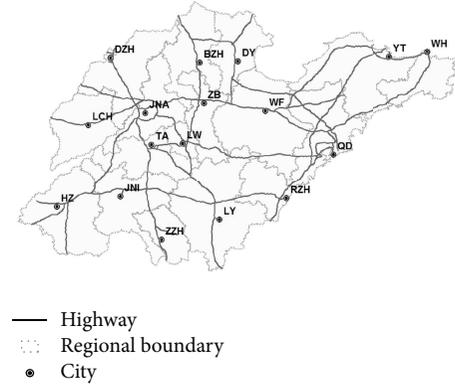


FIGURE 1: A sketch map of city location and highway infrastructure in Shandong province of China.

7 ultimate (32 bits) with 2 GB of RAM. The computational results are presented in Table 2.

When there idealistically exists a situation in which the distribution of terminals and the loaded-semitrailer flows right satisfy all constraints so that any segment on routes has loaded-semitrailer transported, the  $\text{CO}_2$  emissions per ton-kilometer will reach the minimum. According to our assumption, the minimum is  $48.53 \text{ g CO}_2/\text{t-km}$ . Since most instances can hardly satisfy the idealistical requirement, the  $\text{CO}_2$  emissions per ton-kilometer of the LBM will be larger than  $48.53 \text{ g CO}_2/\text{t-km}$ . We calculate the percentage gap by  $\text{gap}_1 = |(\text{LBM} - \text{Min})/\text{Min}| \times 100$  or  $\text{gap}_2 = |(\text{SA} - \text{LBM})/\text{LBM}| \times 100$ . The results show that the average percentage gap between LBM and the theoretical minimum is 4.72%. 72.5% of the 40 test instances have percentage gaps of lower than 5%.

As showed in Table 2, the solution  $\text{CO}_2$  emissions per ton-kilometer obtained by the proposed SA heuristic are close to the lower bound. For some instances, such as RAND5-4, RAND5-8, RAND6-8, RAND7-1, RAND7-2, RAND7-4, RAND7-6, and RAND8-1, the gaps between the heuristic solution and LBM are less than 1%. For all 40 test instances, the largest percentage gap between the heuristic solution and LBM is 14.20%, and the average percentage gap is 2.79%. 87.5% of the 40 instances have percentage gaps of lower than 5%.

It is worth noting that the location of a central depot likely affects the percentage gap. The instances with percentage gaps of over 5% have their central depots located alongside the boundary of the regions.

**5.2. A Realistic Instance.** The main purpose of the realistic instance study is to evaluate the applicability of the developed heuristics for realistic-size problems.

A trucking company in Shandong province of China, simply named SDEXP, is the object of our computational study. SDEXP comprises 17 affiliates distributed in 17 cities of Shandong province (Figure 1). SDEXP employed hundreds of single-unit trucks to transport cargoes before 2007 and had a road freight market share of about 1.25% in Shandong

TABLE 1: Basic characteristics of the 40 test problems.

Problem	Region (km × km)	Freight flow demand (loaded-semitrailer)			Distance between central depot and satellite facility		Distance between any two terminals	
		Total	Average	Variance	Min.	Max.	Average	Variance
RAND5-1	200 × 150	49	2.5	1.0	50	200	126.7	2712.6
RAND5-2	200 × 200	51	2.6	1.0	50	200	133.3	2298.9
RAND5-3	250 × 200	60	2.4	1.1	50	200	163.3	4988.5
RAND5-4	200 × 200	58	2.9	0.8	50	200	133.3	2298.9
RAND5-5	200 × 250	46	2.3	1.1	50	250	176.7	7195.4
RAND5-6	200 × 200	66	2.6	0.9	50	150	143.3	3057.5
RAND5-7	250 × 250	57	2.3	1.0	50	300	186.7	7402.3
RAND5-8	200 × 250	74	2.5	1.0	50	300	173.3	6160.9
RAND5-9	200 × 250	70	2.3	1.0	200	350	183.3	6781.6
RAND5-10	250 × 150	81	2.3	1.2	50	200	123.8	3809.5
RAND6-1	200 × 250	74	2.5	1.0	50	300	173.3	6160.9
RAND6-2	200 × 200	99	2.8	1.3	50	250	147.6	4994.2
RAND6-3	200 × 250	91	2.5	1.2	50	200	161.9	5342.6
RAND6-4	200 × 250	72	2.4	0.9	50	200	138.1	4854.8
RAND6-5	250 × 150	73	2.4	0.9	50	250	142.9	4947.7
RAND6-6	250 × 250	97	2.7	0.9	100	200	185.7	6132.4
RAND6-7	200 × 200	99	2.8	1.3	50	250	147.6	4994.2
RAND6-8	250 × 200	80	2.7	1.0	50	250	161.9	6806.0
RAND6-9	250 × 200	98	2.7	1.1	50	300	142.9	4703.8
RAND6-10	150 × 250	106	2.9	0.7	50	250	138.1	4854.8
RAND7-1	200 × 200	102	2.4	1.1	50	200	132.1	3675.3
RAND7-2	250 × 250	74	2.1	0.8	100	250	176.8	6633.1
RAND7-3	250 × 250	109	2.6	0.9	100	300	162.5	5022.7
RAND7-4	250 × 200	88	2.5	1.3	100	200	155.4	5152.6
RAND7-5	250 × 250	74	2.1	0.8	100	300	176.8	6633.1
RAND7-6	250 × 250	74	2.1	0.8	100	300	176.8	6633.1
RAND7-7	250 × 250	74	2.1	0.8	100	200	176.8	6633.1
RAND7-8	250 × 250	74	2.1	0.8	100	300	176.8	6633.1
RAND7-9	250 × 250	74	2.1	0.8	150	250	176.8	6633.1
RAND7-10	250 × 250	74	2.1	0.8	100	400	176.8	6633.1
RAND8-1	200 × 250	126	2.6	1.2	50	300	141.7	4436.6
RAND8-2	200 × 250	126	2.6	1.2	50	250	141.7	4436.6
RAND8-3	200 × 250	126	2.6	1.2	50	150	141.7	4436.6
RAND8-4	200 × 250	126	2.6	1.2	50	150	141.7	4436.6
RAND8-5	200 × 250	126	2.6	1.2	50	200	141.7	4436.6
RAND8-6	200 × 250	126	2.6	1.2	150	300	141.7	4436.6
RAND8-7	200 × 250	126	2.6	1.2	50	200	141.7	4436.6
RAND8-8	200 × 250	126	2.6	1.2	50	250	141.7	4436.6
RAND8-9	200 × 250	150	3.1	0.8	100	350	163.9	5438.2
RAND8-10	200 × 250	150	3.1	0.8	100	350	163.9	5438.2

Note: the problem is denoted by RAND (number 1)-(number 2) where number 1 is the number of satellite facilities and ranges from 5 to 8, and number 2 is the instance sequence in a same set.

TABLE 2: Computational results for the 40 test problems.

Problem	SA			LBM			gap1 (%)	gap2 (%)
	Percentage of satisfied freight demand (%)	Tractor quantity	CO <sub>2</sub> emissions per ton-kilometer (g/tm)	Percentage of satisfied freight demand (%)	Tractor quantity	CO <sub>2</sub> emissions per ton-kilometer (g/tm)		
RAND5-1	86	13	51.85	86	12	50.39	3.83	2.90
RAND5-2	86	14	54.65	86	12	49.01	0.99	11.51
RAND5-3	87	12	49.99	87	15	49.16	1.30	1.69
RAND5-4	88	16	49.99	86	15	50.27	3.59	0.56
RAND5-5	89	10	57.33	87	12	50.20	3.44	14.20
RAND5-6	88	15	50.56	88	17	49.85	2.72	1.42
RAND5-7	88	13	49.97	86	14	49.13	1.24	1.71
RAND5-8	86	16	50.74	87	19	50.78	4.64	0.08
RAND5-9	89	14	50.77	89	18	51.29	5.69	1.01
RAND5-10	85	14	49.17	85	20	50.14	3.32	1.93
RAND6-1	85	14	48.95	85	19	50.35	3.75	2.78
RAND6-2	90	19	48.92	89	25	49.86	2.74	1.89
RAND6-3	86	18	50.13	86	23	50.79	4.66	1.30
RAND6-4	86	17	49.97	86	18	48.89	0.74	2.21
RAND6-5	89	14	48.71	88	18	49.22	1.42	1.04
RAND6-6	86	16	48.70	86	24	50.25	3.54	3.08
RAND6-7	90	17	48.82	89	25	49.86	2.74	2.09
RAND6-8	88	17	49.82	86	20	49.50	2.00	0.65
RAND6-9	90	17	48.64	89	25	49.95	2.93	2.62
RAND6-10	88	18	48.73	86	27	50.35	3.75	3.22
RAND7-1	90	21	49.84	88	26	50.12	3.28	0.56
RAND7-2	86	17	50.37	87	19	50.59	4.24	0.43
RAND7-3	90	18	49.13	89	27	54.03	11.33	9.07
RAND7-4	86	19	49.72	85	22	50.05	3.13	0.66
RAND7-5	85	16	52.36	85	16	53.08	9.38	1.36
RAND7-6	85	14	51.95	85	12	51.81	6.76	0.27
RAND7-7	85	14	52.08	85	14	53.17	9.56	2.05
RAND7-8	85	16	52.77	85	16	53.79	10.84	1.90
RAND7-9	85	13	54.79	88	13	53.58	10.41	2.26
RAND7-10	85	16	55.69	85	15	53.04	9.29	5.00
RAND8-1	86	21	49.49	86	32	49.74	2.49	0.50
RAND8-2	88	24	52.82	89	13	49.57	2.14	6.56
RAND8-3	89	24	51.35	89	26	50.62	4.31	1.44
RAND8-4	88	24	51.68	89	20	50.17	3.38	3.01
RAND8-5	88	23	50.83	88	24	50.51	4.08	0.63
RAND8-6	89	22	54.88	89	22	53.96	11.19	1.70
RAND8-7	88	22	53.04	89	22	50.33	3.71	5.38
RAND8-8	87	22	52.82	87	26	50.84	4.76	3.89
RAND8-9	87	27	49.99	87	38	52.05	7.25	3.96
RAND8-10	87	26	50.95	87	36	52.61	8.41	3.16

province. Along with the policy on encouraging and popularizing tractor and semitrailer combinations issued by the MOTPRC during the 11th five-year plan period (2006~2010), SDEXP plans to gradually substitute tractor and semitrailer combinations for single-unit trucks.

We abstract the transportation network of SDEXP on a graph, where the nodes denote the cities and the arcs denote road infrastructure connecting every two cities. Suppose any city node can be regarded as a central depot and other city nodes satellite facilities. Table 3 gives the distances between

TABLE 3: The distances between any two cities (km).

	JNA	QD	ZB	ZZH	DY	YT	WF	JNI	TA	WH	RZH	LW	LY	DZH	LCH	BZH	HZ
JNA	0	361	111	245	225	457	209	202	75	517	332	86	250	131	125	159	251
QD	361	0	263	429	275	251	157	448	363	272	175	308	300	469	464	330	557
ZB	111	263	0	313	128	360	112	270	146	419	323	91	240	222	216	83	357
ZZH	245	429	313	0	439	589	424	137	172	649	267	222	116	329	324	372	247
DY	225	275	128	439	0	350	141	398	274	410	359	219	368	335	329	75	470
YT	457	251	360	589	350	0	252	606	504	65	333	449	458	565	559	426	700
WF	209	157	112	424	141	252	0	382	257	314	226	202	249	318	313	179	454
JNI	202	448	270	137	398	606	382	0	130	667	285	180	203	287	156	331	126
TA	75	363	146	172	274	504	257	130	0	566	277	58	195	169	163	207	240
WH	517	272	419	649	410	65	314	667	566	0	395	510	519	626	621	487	762
RZH	332	175	323	267	359	333	226	285	277	395	0	247	137	443	437	383	395
LW	86	308	91	222	219	449	202	180	58	510	247	0	165	225	220	152	290
LY	250	300	240	116	368	458	249	203	195	519	137	165	0	361	355	301	313
DZH	131	469	222	329	335	565	318	287	169	626	443	225	361	0	180	271	338
LCH	125	464	216	324	329	559	313	156	163	621	437	220	355	180	0	263	225
BZH	159	330	83	372	75	426	179	331	207	487	383	152	301	271	263	0	404
HZ	251	557	357	247	470	700	454	126	240	762	395	290	313	338	225	404	0

TABLE 4: The freight flows between two cities (unit: loaded-semitrailer).

To	From																
	JNA	QD	ZB	ZZH	DY	YT	WF	JNI	TA	WH	RZH	LW	LY	DZH	LCH	BZH	HZ
JNA	0	1	2	1	2	1	3	1	0	0	0	0	1	0	1	2	1
QD	1	0	6	0	4	6	17	0	1	2	3	1	2	1	1	1	0
ZB	1	4	0	1	2	2	3	1	1	1	0	0	2	3	3	0	1
ZZH	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
DY	2	3	2	0	0	1	1	0	1	0	0	0	1	1	1	0	0
YT	1	6	2	0	2	0	4	0	0	1	1	0	1	0	0	0	0
WF	3	14	3	0	1	4	0	0	1	1	1	1	2	1	1	3	0
JNI	1	0	1	0	0	0	0	0	0	0	0	0	1	0	2	0	0
TA	0	1	1	1	1	0	1	0	0	0	0	0	2	2	3	1	1
WH	0	2	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0
RZH	0	3	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
LW	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	1	0
LY	1	2	3	0	1	1	2	1	2	0	0	1	0	0	0	1	0
DZH	0	0	3	0	1	0	1	1	2	0	0	0	0	0	2	1	0
LCH	1	1	3	0	1	0	1	2	2	0	0	1	0	2	0	1	1
BZH	2	1	0	0	0	0	4	0	1	0	0	1	1	1	1	0	0
HZ	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0

every two cities. SDEXP’s expected freight flows between any two cities per day are given in Table 4.

The tractor-semitrailer combination adopted in the small-scale instances is used in the realistic instance. Suppose that the loading factor of loaded-semitrailer is 50%; the velocity is 80 km/h; and the percentage of satisfied freight flow demand is not less than 80%. The fuel consumption is 17 liters diesel per 100 kilometers for a tractor that runs alone and 40 liters diesel per 100 kilometers for a combination runs. The distance span is affected by drivers’ on-duty time. According to SDEXP experience, a driver’s average on-duty time is 8.5 hours per day. A tractor with two drivers can

work consecutively for not more than 17.0 hours in a 24-consecutive-hour period. The attach/detach time at satellite facilities is 2/3 hour, and the residence time in the central depot is 2 hours.  $\rho_1 = 0.9$ .  $\rho_2 = 1.1$ .

Supposing any one of the 17 cities the candidate central depot city for SDEXP, there are 17 scenarios classified by the location of the central depot. The satisfactory solutions of the 17 scenarios are in Table 5. When cities located in different spatial zone and having various freight flows are, respectively, regarded as the central depot, the performance of the satisfactory solution varies obviously. We use “Tractor quantity;” “Percentage of fuel consumption for tractor

TABLE 5: The results of the 17 scenarios of the realistic instance.

The central depot city	Tractor quantity	Percentage of satisfied freight demand (%)	Average number of loaded-semitrailer on a route	Average distance span (km)	Percentage of fuel consumption for tractor running alone (%)	CO <sub>2</sub> emissions per ton-kilometer (g/tm)
JNA	52	82	4.20	1068	7.20	78.45
JNA*	62	80	3.44	1000	9.04	80.04
QD	44	81	4.93	1102	1.52	73.92
QD*	52	80	4.12	1036	3.75	75.64
ZB	52	81	4.12	1068	3.61	75.53
ZB*	55	81	3.93	1033	4.54	76.27
ZZH	76	80	2.82	1288	28.60	101.96
DY	56	80	3.82	1077	7.34	78.57
DY*	59	80	3.63	1045	8.39	79.47
YT	99	80	2.16	1082	32.66	108.11
WF	49	81	4.43	1047	1.48	73.89
WF*	51	80	4.20	1017	2.47	74.65
JNI	77	80	2.77	1159	27.06	99.81
TA	62	81	3.46	974	8.68	79.72
TA*	64	82	3.39	1010	10.05	80.93
WH	102	80	2.09	1382	43.02	127.76
RZH	72	81	2.99	1074	18.42	89.24
LW	61	81	3.52	1003	8.29	79.38
LW*	63	81	3.41	1026	10.03	80.91
LY	65	82	3.35	1075	15.02	85.67
LY*	67	81	3.21	1065	16.09	86.76
DZH	68	82	3.19	1121	19.07	89.95
LCH	64	80	3.33	1093	15.71	86.37
BZH	57	81	3.77	1028	6.61	77.95
BZH*	59	81	3.64	1023	7.82	78.98
HZ	77	80	2.77	1328	30.78	105.17

\*The results without clone operator.

running alone,” and “CO<sub>2</sub> emissions per ton-kilometer” to analyze the performance of the solutions.

At an approximative level of the percentage of satisfied freight demand, the needed quantity of tractors is about 64 for different central depot location except YT and WH. The average quantity of tractors is around 1/4 of total freight demand.

There exists a relationship between the quantity of needed tractors and the average loaded-semitrailer number of a route. Since the number of loaded-semitrailers is decided by the percentage of satisfied freight demand, the less the tractor number (or the route number), the more the loaded semitrailer on a route. When the number of loaded-semitrailers and the distance span are decided, the more the tractor number, the more the percentage of fuel consumption for tractor running alone. The results are accordant with the above two relationships (Figure 2). The SA heuristics is stable to solve the realistic instance.

According to the methodology and factors developed by the Intergovernmental Panel on Climate Change (IPCC), CO<sub>2</sub> emissions are in direct proportion to fuel consumption,

so CO<sub>2</sub> emissions of the solutions can express the variable cost of transportation. Escobar et al. [43] and Hashemi and Seifi [39] pointed that the transportation costs are often influenced by the decision of locating a depot and vice versa. Our results show a similar conclusion. The solutions for various central depot cities have different CO<sub>2</sub> emissions per ton-kilometer. Central depot cities located near the center of the research spatial scope (e.g., TA, LW, JNA, and ZB) have relatively good solutions. Besides, central depot cities located along Jinan-Qingdao Highway (JNA-ZB-WF-QD line) have good solutions that include more loaded-semitrailers on a route and a low level of CO<sub>2</sub> emissions per ton-kilometer. In fact, the transportation economic belt along Jinan-Qingdao Highway contributed around 40% of Shandong GDP and over 30% of Shandong road freight volume in the recent 5 years. It is implied that the CO<sub>2</sub> emissions per ton-kilometer of the solutions for different central depot cities are affected not only by central depot locations but also by transportation flows from economic relations.

There are estimation results of CO<sub>2</sub> emissions per ton-kilometer for various countries. For example, Ierland et al.

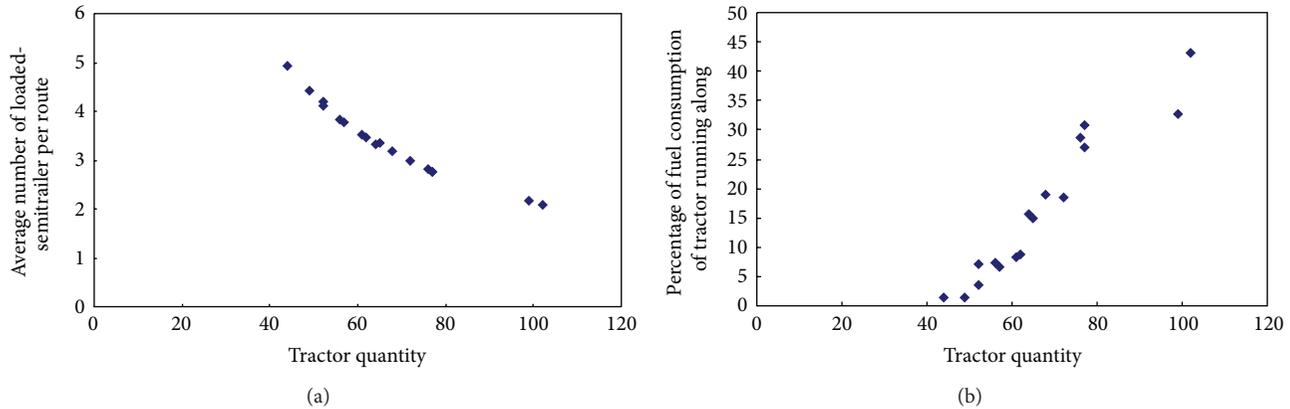


FIGURE 2: In the SA results, (a) the relationship between the tractor quantity and the average loaded-semitrailer number of a route; (b) the relationship between the tractor quantity and the percentage of fuel consumption for tractor running alone.

[13] noted that the  $\text{CO}_2$  emission factor for trucks was  $155 \text{ g CO}_2/\text{t-km}$  in the Netherlands; European Environment Agency [44] noted that the average  $\text{CO}_2$  emissions were  $62\sim 110 \text{ g CO}_2/\text{t-km}$  for road transportation in EU Member States; Li et al. [3] noted that the average  $\text{CO}_2$  emissions fluctuated between  $100 \text{ g CO}_2/\text{t-km}$  and  $132 \text{ g CO}_2/\text{t-km}$  from 1985 to 2007 in China. We have investigated some point-to-point haulages of SDEXP in 2009 and found that the  $\text{CO}_2$  emission factor ranged from  $100 \text{ g CO}_2/\text{t-km}$  to  $180 \text{ g CO}_2/\text{t-km}$ , and the average was  $135 \text{ g CO}_2/\text{t-km}$ . Our realistic-instance study shows that the vehicle scheduling provided by the TSRP is promising to reduce road freight transportation  $\text{CO}_2$  emissions.

The results are helpful and meaningful for SDEXP to select the location of the central depot, to configure tractors and the routes, and to reduce  $\text{CO}_2$  emissions.

## 6. Conclusions and Future Work

This paper discussed the tractor and semitrailer routing problem and its mitigation effect of  $\text{CO}_2$  emissions, which are promising with applications in multilevel freight distribution systems and full truckload to road freight-related  $\text{CO}_2$  emissions reduction. A mathematical programming model with the objective of minimizing  $\text{CO}_2$  emissions per ton-kilometer is presented for the tractor and semitrailer routing problem with full truckloads between any two terminals of the network. The SA heuristic is put forward to solve this problem of a realistic size. To validate the proposed heuristic, a lower bound is designed. The heuristic algorithm is tested on different types of problems. The results show that the proposed heuristic provides high-quality solutions in a reasonable computing time. The impact of the central depot location and the freight flow distribution on the solution quality is also explored. In conclusion, the proposed algorithm can provide robust solutions.

For future research, it would be interesting to test the effectiveness and efficiency of the proposed TSRP model and its solution approach on various practical examples. Some efficient heuristics for the TSRP may also be proposed.

Besides, attentions can be focused on the extension of the TSRP, for example, TSRP with time windows and TSRP with vehicle routing of other levels of freight distribution system, which are properly of critical importance to the practical viability.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

This work was supported by Research Grant from the National Natural Science Foundation of China (no. 71202016). This support is gratefully acknowledged.

## References

- [1] GFEI, "Improving vehicle fuel economy in the ASEAN region," 2010, [http://www.globalfueleconomy.org/Documents/Publications/wp1\\_asean\\_fuel\\_economy.pdf](http://www.globalfueleconomy.org/Documents/Publications/wp1_asean_fuel_economy.pdf).
- [2] A. C. McKinnon and M. I. Piecyk, "Measurement of  $\text{CO}_2$  emissions from road freight transport: a review of UK experience," *Energy Policy*, vol. 37, no. 10, pp. 3733–3742, 2009.
- [3] H. Q. Li, Y. Lu, J. Zhang, and T. Y. Wang, "Trends in road freight transportation carbon dioxide emissions and policies in China," *Energy Policy*, vol. 57, pp. 99–106, 2013.
- [4] D. Y. Lin and K. H. Ng, "The impact of collaborative backhaul routing on carbon reduction in the freight industry," *Transportation Research Part D*, vol. 17, no. 8, pp. 626–628, 2012.
- [5] R. Baños, J. Ortega, C. Gil, A. Fernández, and F. Toro, "A simulated annealing-based parallel multi-objective approach to vehicle routing problems with time windows," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1696–1707, 2013.
- [6] X. J. Zhang, Z. S. He, Z. He, and L. Rayman-Bacchus, "Probability routing strategy for scale-free networks," *Physica A*, vol. 392, no. 4, pp. 953–958, 2013.
- [7] G. Laporte, "Fifty years of vehicle routing," *Transportation Science*, vol. 43, no. 4, pp. 408–416, 2009.

- [8] R. Baldacci, P. Toth, and D. Vigo, "Recent advances in vehicle routing exact algorithms," *4OR*, vol. 5, no. 4, pp. 269–298, 2007.
- [9] Y. Kuo, "Using simulated annealing to minimize fuel consumption for the time-dependent vehicle routing problem," *Computers & Industrial Engineering*, vol. 59, no. 1, pp. 157–165, 2010.
- [10] L. Pradenas, B. Oportus, and V. Parada, "Mitigation of greenhouse gas emissions in vehicle routing problems with backhauling," *Expert Systems with Applications*, vol. 40, no. 8, pp. 2985–2991, 2013.
- [11] M. Drexl, "Applications of the vehicle routing problem with trailers and transshipments," *European Journal of Operational Research*, 2012.
- [12] H. Q. Li and Y. Lu, "Tractor and semi-trailer transportation's effect on the reduction of CO<sub>2</sub> emission in China," in *Proceeding of the 7th Advanced Forum on Transportation of China (AFTC '11)*, B. H. Mao, S. J. Yang, and L. L. Yang, Eds., pp. 232–237, The Institution of Engineering and Technology: Systems Science and Transportation Development, Beijing, China, 2011.
- [13] E. V. Ierland, C. Graveland, and R. Huiberts, "An environmental economic analysis of the new rail link to european main port Rotterdam," *Transportation Research Part D*, vol. 5, no. 3, pp. 197–209, 2000.
- [14] V. C. Hemmelmayr, J. F. Cordeau, and T. G. Crainic, "An adaptive large neighborhood search heuristic for Two-Echelon vehicle routing problems arising in city logistics," *Computers and Operations Research*, vol. 39, no. 12, pp. 3215–3228, 2012.
- [15] P. C. Pop, I. Kara, and A. H. Marc, "New mathematical models of the generalized vehicle routing problem and extensions," *Applied Mathematical Modelling*, vol. 36, no. 1, pp. 97–107, 2012.
- [16] C. Contardo, V. Hemmelmayr, and T. G. Crainic, "Lower and upper bounds for the two-echelon capacitated location-routing problem," *Computers & Operations Research*, vol. 39, no. 12, pp. 3185–3199, 2012.
- [17] I. M. Chao, "A tabu search method for the truck and trailer routing problem," *Computers and Operations Research*, vol. 29, no. 1, pp. 33–51, 2002.
- [18] F. Semet and E. Taillard, "Solving real-life vehicle routing problems efficiently using tabu search," *Annals of Operations Research*, vol. 41, no. 4, pp. 469–488, 1993.
- [19] J. C. Gerdessen, "Vehicle routing problem with trailers," *European Journal of Operational Research*, vol. 93, no. 1, pp. 135–147, 1996.
- [20] M. Caramia and F. Guerriero, "A milk collection problem with incompatibility constraints," *Interfaces*, vol. 40, no. 2, pp. 130–143, 2010.
- [21] S. Scheuerer, "A tabu search heuristic for the truck and trailer routing problem," *Computers and Operations Research*, vol. 33, no. 4, pp. 894–909, 2006.
- [22] K. C. Tan, Y. H. Chew, and L. H. Lee, "A hybrid multi-objective evolutionary algorithm for solving truck and trailer vehicle routing problems," *European Journal of Operational Research*, vol. 172, no. 3, pp. 855–885, 2006.
- [23] S. W. Lin, V. F. Yu, and S. Y. Chou, "Solving the truck and trailer routing problem based on a simulated annealing heuristic," *Computers & Operations Research*, vol. 36, no. 5, pp. 1683–1692, 2009.
- [24] J. G. Villegas, C. Prins, C. Prodhon, A. L. Medaglia, and N. Velasco, "A GRASP with evolutionary path relinking for the truck and trailer routing problem," *Computers & Operations Research*, vol. 38, no. 9, pp. 1319–1334, 2011.
- [25] J. G. Villegas, C. Prins, C. Prodhon, A. L. Medaglia, and N. Velasco, "GRASP/VND and multi-start evolutionary local search for the single truck and trailer routing problem with satellite depots," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 5, pp. 780–794, 2010.
- [26] S. W. Lin, V. F. Yu, and S. Y. Chou, "A note on the truck and trailer routing problem," *Expert Systems with Applications*, vol. 37, no. 1, pp. 899–903, 2010.
- [27] S. W. Lin, V. F. Yu, and C. C. Lu, "A simulated annealing heuristic for the truck and trailer routing problem with time windows," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15244–15252, 2011.
- [28] U. Derigs, M. Pullmann, and U. Vogel, "Truck and trailer routing—problems, heuristics and computational experience," *Computers & Operations Research*, vol. 40, no. 2, pp. 536–546, 2013.
- [29] V. Pureza, R. Morabito, and M. Reimann, "Vehicle routing with multiple deliverymen: modeling and heuristic approaches for the VRPTW," *European Journal of Operational Research*, vol. 218, no. 3, pp. 636–647, 2012.
- [30] L. Bodin, A. Mingozzi, R. Baldacci, and M. Ball, "The rollon-rolloff vehicle routing problem," *Transportation Science*, vol. 34, no. 3, pp. 271–288, 2000.
- [31] U. Derigs, M. Pullmann, and U. Vogel, "A short note on applying a simple LS/LNS-based metaheuristic to the rollon-rolloff vehicle routing problem," *Computers & Operations Research*, vol. 40, no. 3, pp. 867–872, 2013.
- [32] J. Wy, B.-I. Kim, and S. Kim, "The rollon-rolloff waste collection vehicle routing problem with time windows," *European Journal of Operational Research*, vol. 224, no. 3, pp. 466–476, 2013.
- [33] R. Baldacci, L. Bodin, and A. Mingozzi, "The multiple disposal facilities and multiple inventory locations rollon-rolloff vehicle routing problem," *Computers and Operations Research*, vol. 33, no. 9, pp. 2667–2702, 2006.
- [34] R. W. Hall and V. C. Sabnani, "Control of vehicle dispatching on a cyclic route serving trucking terminals," *Transportation Research Part A*, vol. 36, no. 3, pp. 257–276, 2002.
- [35] P. Francis, G. M. Zhang, and K. Smilowitz, "Improved modeling and solution methods for the multi-resource routing problem," *European Journal of Operational Research*, vol. 180, no. 3, pp. 1045–1059, 2007.
- [36] Y. R. Cheng, B. Liang, and M. H. Zhou, "Optimization for vehicle scheduling in iron and steel works based on semi-trailer swap transport," *Journal of Central South University of Technology*, vol. 17, no. 4, pp. 873–879, 2010.
- [37] U. Derigs, R. Kurowsky, and U. Vogel, "Solving a real-world vehicle routing problem with multiple use of tractors and trailers and EU-regulations for drivers arising in air cargo road feeder services," *European Journal of Operational Research*, vol. 213, no. 1, pp. 309–319, 2011.
- [38] H. Q. Li, Y. Lu, J. Zhang, and T. Y. Wang, "Solving the tractor and semi-trailer routing problem based on a heuristic approach," *Mathematical Problems in Engineering*, vol. 2012, Article ID 182584, 12 pages, 2012.
- [39] D. S. H. Hashemi and A. Seifi, "Lower and upper bounds for location-arc routing problems with vehicle capacity constraints," *European Journal of Operational Research*, vol. 224, no. 1, pp. 189–208, 2013.
- [40] S. Almoustafa, S. Hanafi, and N. Mladenović, "New exact method for large asymmetric distance-constrained vehicle routing problem," *European Journal of Operational Research*, vol. 226, no. 3, pp. 386–394, 2013.

- [41] O. Braysy and M. Gendreau, “Vehicle routing problem with time windows, part I: route construction and local search algorithms,” *Transportation Science*, vol. 39, no. 1, pp. 104–118, 2005.
- [42] L. C. Coelho, J. F. Cordeau, and G. Laporte, “Consistency in multi-vehicle inventory-routing,” *Transportation Research Part C*, vol. 24, pp. 270–287, 2012.
- [43] J. W. Escobar, R. Linfati, and P. Toth, “A two-phase hybrid heuristic algorithm for the capacitated location-routing problem,” *Computers & Operations Research*, vol. 40, no. 1, pp. 70–79, 2013.
- [44] European Environment Agency, “Transport at a crossroads—TERM 2008: indicators tracking transport and environment in the European Union,” Tech. Rep., Office for Official Publications of the European Communities, Luxembourg, 2009.

## Research Article

# Comparison of Electric Vehicle's Energy Consumption Factors for Different Road Types

Enjian Yao,<sup>1,2</sup> Zhiqiang Yang,<sup>1,2</sup> Yuanyuan Song,<sup>2</sup> and Ting Zuo<sup>2</sup>

<sup>1</sup> MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup> School of Traffic and Transportation, Beijing Jiaotong University, Haidian District, Beijing 100044, China

Correspondence should be addressed to Zhiqiang Yang; 12120938@bjtu.edu.cn

Received 30 September 2013; Revised 23 November 2013; Accepted 5 December 2013

Academic Editor: Huimin Niu

Copyright © 2013 Enjian Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Energy-optimal route planning for electric vehicle (EV) is highly required for the wide-spread use of EV, which is hindered by limited battery capacity and relative short cruising range. Obtaining the cost for each link (i.e., link energy consumption) in road networks plays a key role in energy-optimal route planning process. The link energy consumption depends mainly on energy consumption factor, which is related to not only vehicle speed but also road type. This study aims to analyze the difference of EV's energy consumption factors for different road types. According to the floating car data (FCD) collected from the road network in Beijing, the vehicle specific power (VSP) distributions under different average travel speeds for different road types are analyzed firstly, and then the EV's energy consumption rates under different VSP-Bins are calculated. By using VSP as an intermediate variable, EV's energy consumption factor models for different road types are established and the difference of EV's energy consumption factors is analyzed. The results show that road type-based energy consumption factor should be used in EV's energy-optimal route planning process.

## 1. Introduction

Recently, to cope with the problem of pollutant emissions and energy consumption caused by gasoline and diesel powered vehicles, increasing attention has been paid to EVs due to the advantages of zero-emission during use, low noise, and high energy efficiency. However, the relative short cruising range has become one of the main obstacles to the development and wide-spread use of EVs [1]. Some intelligent transportation systems (ITS) solutions such as energy-efficient route planning and charging facilities guidance systems are proposed to help EV drivers to optimize travel route and find charging station timely, which are expected to alleviate the restrictions resulted from the relative short cruising range of EVs. Estimating EV's energy consumption accurately is a prerequisite in route planning and navigation systems [2–4]; therefore it is of great significance to estimate EV's energy consumption.

Many researches on EV's energy consumption estimation are based on ideal running status, whose results cannot reflect

the effect of the actual vehicle running status. Physically based methods, which take into account vehicle driving parameters (i.e., average travel speed), are usually used in modeling fuel consumption of gasoline and diesel powered vehicles [5]. In these models, VSP is introduced as an intermediate variable for the ability to build the relationship between energy consumption and vehicle microscopic driving parameters (i.e., instantaneous speed and acceleration). However, due to the difficulty in collecting the data for reflecting urban roads' real driving situation, existing researches scarcely take the difference of VSP distributions for different road types into consideration. A great many practices have proved that energy consumption varies greatly in different road types even when vehicles travel at a fixed average speed because the frequency and range of acceleration and deceleration are always different. Therefore, in order to estimate the EV's energy consumption for different road types with high accuracy, this paper attempts to establish energy consumption factor models for different road types and make a comparison among different road types.

In the research field of EV's energy consumption models, energy consumption is usually equal to the integral of power of battery times the time, but this method fails to reflect the influence of running status on energy consumption. Some studies have already been done on how the influence of running status on energy consumption is presented [6]. For macroscopic models, vehicle average travel speed is always used as the independent variable and the speed correction factors are used to adjust the energy consumption factors for specific facility types [7]. However, it is questionable whether this approach can reflect the effects of various operation modes, such as frequent acceleration, deceleration, and idling on urban roads [8–11].

Motivated by the challenge associated with the preceding problem, related researches on the models which estimate EV's energy consumption with microscopic driving parameters as input have been conducted extensively. In the development of microscopic models, some researchers incorporate the parameters of instantaneous speed and acceleration. Yao et al. [12] establish a set of electricity consumption rate models and gasoline consumption rate models with instantaneous speed and acceleration as input parameters, to explore the relationship between the EV penetration and the reduction of energy consumption. Hansen et al. [13] and Zachariadis and Samaras [14] develop speed/acceleration matrices that provide fuel consumption per unit time for each combination of instantaneous speed and acceleration. In the VT-Micro model, Ahn et al. [15] use the combination of linear, quadratic, and cubic speed and acceleration terms for the model development by using the regression method. However, it is not realistic to expect that such a substantial number of microscopic driving parameters are always available in most current traffic information systems.

Other researches utilize VSP in the modeling of energy consumption. The VSP (kw/t), which is defined as the instantaneous power per unit mass of the vehicle [16], is verified to have direct physical interpretation of and strong statistical correlations with energy consumption [17]. After the first application by Jiménez-Palacios [16], VSP has been further developed and applied by other researchers in the area of fuel consumption modeling [18]. By analyzing the second-by-second data of fuel vehicle activities from floating cars and the on-road fuel consumption, Song and Yu [19] point out that VSP distribution in a traffic network determines the fuel consumption per unit time, and further, the combination of VSP distribution and average travel speed determines the fuel consumption per unit distance. In their studies, an approach for evaluating the fuel efficiency of road traffic is proposed on the basis of the characterization of VSP distribution and average travel speed from second-by-second data. Yao and Song [20] establish fuel consumption and emission models for light-duty gasoline vehicle, mid-duty diesel vehicle, and heavy-duty diesel vehicle separately based on vehicle average travel speed by using VSP as an intermediate variable. Liu and Yu [21] collect a mass of floating car data, analyze the VSP distributions for different road types in Beijing, respectively, and then provide

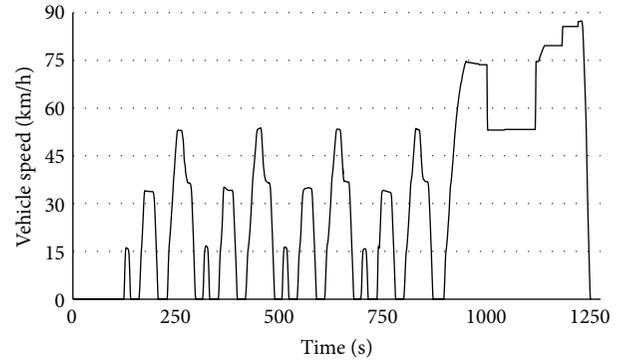


FIGURE 1: Driving cycle of data collection.

speed correction factors for proposed fuel consumption and emission models. Although the researches about fuel consumption are quite mature, there are almost no studies on the EV's energy consumption models for different road types.

The remaining part of this paper is organized as follows. The data utilized in modeling are introduced in Section 2.1 and a method of binning VSP is introduced in Section 2.2. The procedure of modeling EV's energy consumption is described in Section 2.3. In Section 3, a comparison of the VSP-Bin distributions for different road types is given, and then EV's energy consumption factor models for different road types are provided, respectively, and compared. The conclusions are given in Section 4 finally.

## 2. Energy Consumption Factor Models

This paper attempts to establish EV's energy consumption models for different road types, respectively. Urban roads in this paper are classified into express way, arterial road, secondary road, and branch. The influence of acceleration and deceleration on EV's energy consumption is reflected by comparing the VSP-Bin distributions for different road types. The energy consumption factor model for each road type is established, respectively, in which, instead of microscopic driving parameters, average travel speed, which is available in most current traffic information systems, is used as independent variable and VSP is used as intermediate variable.

*2.1. Data Source.* The data used in calculating the energy consumption rates under different VSP-Bins are collected from chassis dynamometer test of a light-duty EV based on the New European Driving Cycle (NEDC) as shown in Figure 1, in which instantaneous speed, acceleration, and battery output power are provided. The data used in analyzing VSP-Bin distributions for different road types are the floating car data collected from the road network in Beijing, in which speed and acceleration per second are provided. Figure 2 shows the test road network in Beijing, while Table 1 and Figure 3 summarized the data sample sizes and the speed distribution for different road types, respectively.

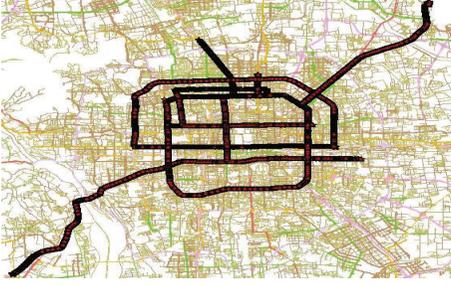


FIGURE 2: The test road network in Beijing.

TABLE 1: The data sample sizes for different road types.

Road type	Sample size
Express way	11378
Arterial road	4674
Secondary road	8853
Branch	5373

2.2. *Vehicle Specific Power.* In this paper, simplified formula is used for the VSP calculation of a light-duty vehicle for various road types, as shown in the following [17]:

$$\text{VSP} = v \times (1.1 \times a + 0.132) + 0.000302 \times v^3, \quad (1)$$

where  $v$  refers to the vehicle speed, m/s, and  $a$  refers to the vehicle acceleration,  $\text{m/s}^2$ .

The VSP for each record from the chassis dynamometer test is calculated according to formula (1), and these VSP values are discrete. The characteristics of EV's energy consumption change greatly under different running status. In this paper, VSP is divided into different VSP-Bins by an interval of 1 kw/t as described in formula (2), so any running status could be modeled on the basis of calculating the time spent in each VSP-Bin [22]. By examining the VSP values of all 48976 records, it is found that 48959 (99%) records fall into  $-30$  to  $30$  kw/t, which means the VSP range from  $-30$  to  $30$  kw/t covers most the running status in the chassis dynamometer test. In particular,  $\text{VSP} = 0$  is set as a single VSP-Bin in this paper because the accuracy of the energy consumption rate diminishes if  $\text{VSP} = 0$  belongs to the bin  $(0, 1]$  or  $[-1, 0)$  as the size of the sample is relatively large when  $\text{VSP} = 0$ :

VSP-Bin

$$= \begin{cases} n, & \forall: \text{VSP} \in [n, n+1) \quad (n = -30, -29, \dots, -1) \\ 0, & \text{VSP} = 0 \\ n, & \forall: \text{VSP} \in (n-1, n] \quad (n = 1, 2, \dots, 30). \end{cases} \quad (2)$$

2.3. *Modeling Methodology.* After binning the VSP, the average energy consumption rate for each VSP-Bin is calculated. In order to build the connection between energy consumption rate and average speed, the second-by-second floating car data collected from the urban road in Beijing are divided into traveling segments by the interval of 60 s, and each segment is characterized with its average travel speed. Then,

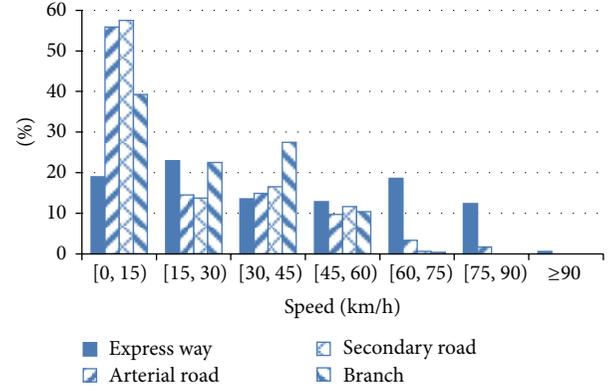


FIGURE 3: The speed distribution for different road types.

by dividing the average travel speed of EVs into different bins by the step of 3.6 km/h, all the travelling segments are classified into different speed bins according to their average travel speeds, and the VSP-Bin distributions under all average travel speed bins for different road types are obtained. The average energy consumption rate under each average travel speed bin is estimated as

$$\overline{\text{EC}}_i = \sum_j \text{EC}_j \times \varphi_{ij}, \quad (3)$$

where  $\overline{\text{EC}}_i$  refers to the average energy consumption rate for the  $i$ th average travel speed bin, kw;  $\text{EC}_j$  refers to the energy consumption rate for the  $j$ th VSP-Bin, kw;  $\varphi_{ij}$  refers to the percentage of the  $j$ th VSP-Bin for the  $i$ th average travel speed bin.

Once the energy consumption rate under each average travel speed is obtained, the energy consumption factors for each road type then can be estimated as

$$V_i = \frac{\sum_{k=1}^n D_{ik}}{\sum_{k=1}^n T_{ik}}, \quad (4)$$

$$\text{EF}_i = \frac{\overline{\text{EC}}_i \times \sum_{k=1}^n T_{ik}}{\sum_{k=1}^n D_{ik}},$$

where  $T_{ik}$  refers to the travel time of EV spent in the  $k$ th segment of the  $i$ th average travel speed bin, h;  $D_{ik}$  refers to the mileage of EV during the  $k$ th segment of the  $i$ th average travel speed bin, km;  $V_i$  refers to the vehicle average travel speed for the  $i$ th average travel speed bin, km/h;  $\text{EF}_i$  refers to the EV's energy consumption factor for the  $i$ th average travel speed bin, kwh/km;  $n$  refers to the number of driving segments.

EV's energy consumption factor models for different road types are established using the multiple linear regression approach, and an energy consumption factor is represented by a function of average travel speed [23] as follows:

$$\text{EF} = \frac{a}{v} + bv + cv^2 + d, \quad (5)$$

where  $v$  refers to the average travel speed of EV, km/h;  $\text{EF}$  refers to EV's energy consumption factor, kwh/km;  $a$ ,  $b$ ,  $c$ , and  $d$  are the coefficients.

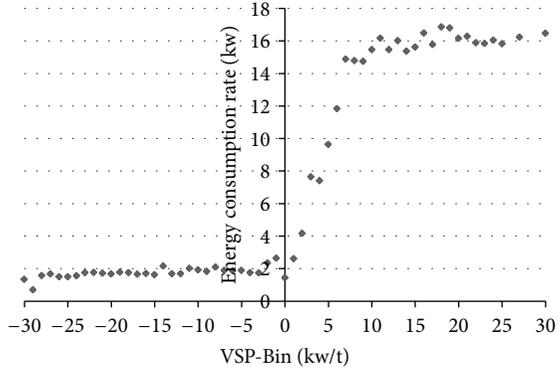


FIGURE 4: Statistics of the energy consumption rates under different VSP-Bins.

### 3. Results

The statistics of the energy consumption rates under different VSP-Bins are shown in Figure 4. The energy consumption rate is low and stable when the VSP-Bin value is negative and increases rapidly as VSP-Bin increases when  $0 < \text{VSP-Bin} < 10$ . However, the rate of increase of the energy consumption rate slows down when  $\text{VSP-Bin} > 10$  and approaches zero finally.

On the basis of (2), the VSP-Bin distribution of each average travel speed bin for each road type is calculated. The measured highest average travel speed bin for arterial road, secondary road, and branch is 50.4–54 km/h and express way has the highest average travel speed bin of 86.4–90 km/h. As an example of illustration, eight average travel speed bins of 0~3.6, 7.2~10.8, 14.4~18, 21.6~25.2, 28.8~32.4, 36~39.6, 43.2~46.8, and 50.4~54 km/h are selected and plotted for demonstrating the general characteristics and variation patterns of VSP-Bin distributions for different road types, as shown in Figure 5. Several characteristics and variation patterns can be observed. (1) For each road type, in the low speed bins, a large percentage of VSP appears in the bin of 0 kw/t. Typical running status of vehicle is idling or queuing around an intersection when  $\text{VSP-Bin} = 0$ . (2) Compared to other road types, the percentage of  $\text{VSP-Bin} = 0$  for arterial road is much lower when average travel speed bin is 0–3.6 (Figure 5(a)), mainly owing to the more frequent traffic congestion on arterial road of the road network in Beijing. (3) With the increase of the average travel speed, the VSP-Bin with the highest percentage increases monotonically, and the highest percentage decreases gradually. (4) The graph with the average travel speed more than 21.6–25.2 km/h shows characteristics similar to a normal distribution.

Based on the average travel speeds and corresponding measured EV's energy consumption factors calculated by (4), the coefficients of (5) for different road types are calibrated, and the results are shown in Table 2. All the model coefficients are significantly not zero (absolute t-values exceed 1.96), which indicate that the estimated result are validated. The adjusted  $R$ -squared for each model is great to indicate a good fit for the EV's energy consumption factor estimation.

TABLE 2: Results of parameters calibration for energy consumption factor models.

Road type	Express way	Arterial road	Secondary road	Branch
Constant				
Coefficient	0.247	-0.179	0.21	0.208
( $t$ -value)	(15.271)	(-2.555)	(16.506)	(24.417)
$\nu$				
Coefficient	-0.004	0.004	-0.001	-0.002
( $t$ -value)	(-5.471)	(2.051)	(-4.067)	(-6.779)
$\nu^{-1}$				
Coefficient	1.520	5.492	1.531	1.553
( $t$ -value)	(48.387)	(56.861)	(61.547)	(210.891)
$\nu^2$				
Coefficient	$2.992E-5$	—	—	—
( $t$ -value)	(4.112)	—	—	—
Adjusted $R^2$	0.996	0.997	0.998	0.999

Note:  $\nu$  is the average travel speed, km/h.

TABLE 3: Energy consumption factor models.

Road type	Calculation equation
Express way	$EF = 0.247 + 1.520/\nu - 0.004\nu + 2.992 \times 10^{-5}\nu^2$
Arterial road	$EF = -0.179 + 0.004\nu + 5.492/\nu$
Secondary road	$EF = 0.21 - 0.001\nu + 1.531/\nu$
Branch	$EF = 0.208 - 0.002\nu + 1.553/\nu$

Note: EF is the energy consumption factor, kwh/km, and  $\nu$  is the average travel speed, km/h.

The energy consumption factor models for different road types are summarized in Table 3.

From the comparison of energy consumption factors under different average travel speeds for different road types (shown in Figure 6), several characteristics and variation patterns can be observed. (1) The EV's energy consumption factors for express way, secondary road, and branch obey a similar change rule. However, for arterial road, EV's energy consumption factor decreases with the increase of average travel speed and then turns to increase when average travel speed reaches a specified value about 37.1 km/h. It means that EV travelling on arterial road is the most energy-efficient situation when it travels at the speed of 37.1 km/h. (2) The energy consumption factor for arterial road is greatly higher than the energy consumption factors for the other three road types when the average travel speed is less than 12 km/h and lower than the other three road types when the average travel speed exceeds 12 km/h. According to above analysis, a conclusion could be drawn that the estimation of EV's energy consumption will be more precise with the road type-based energy consumption factor used in link energy cost calculation in EV's energy-optimal route planning process.

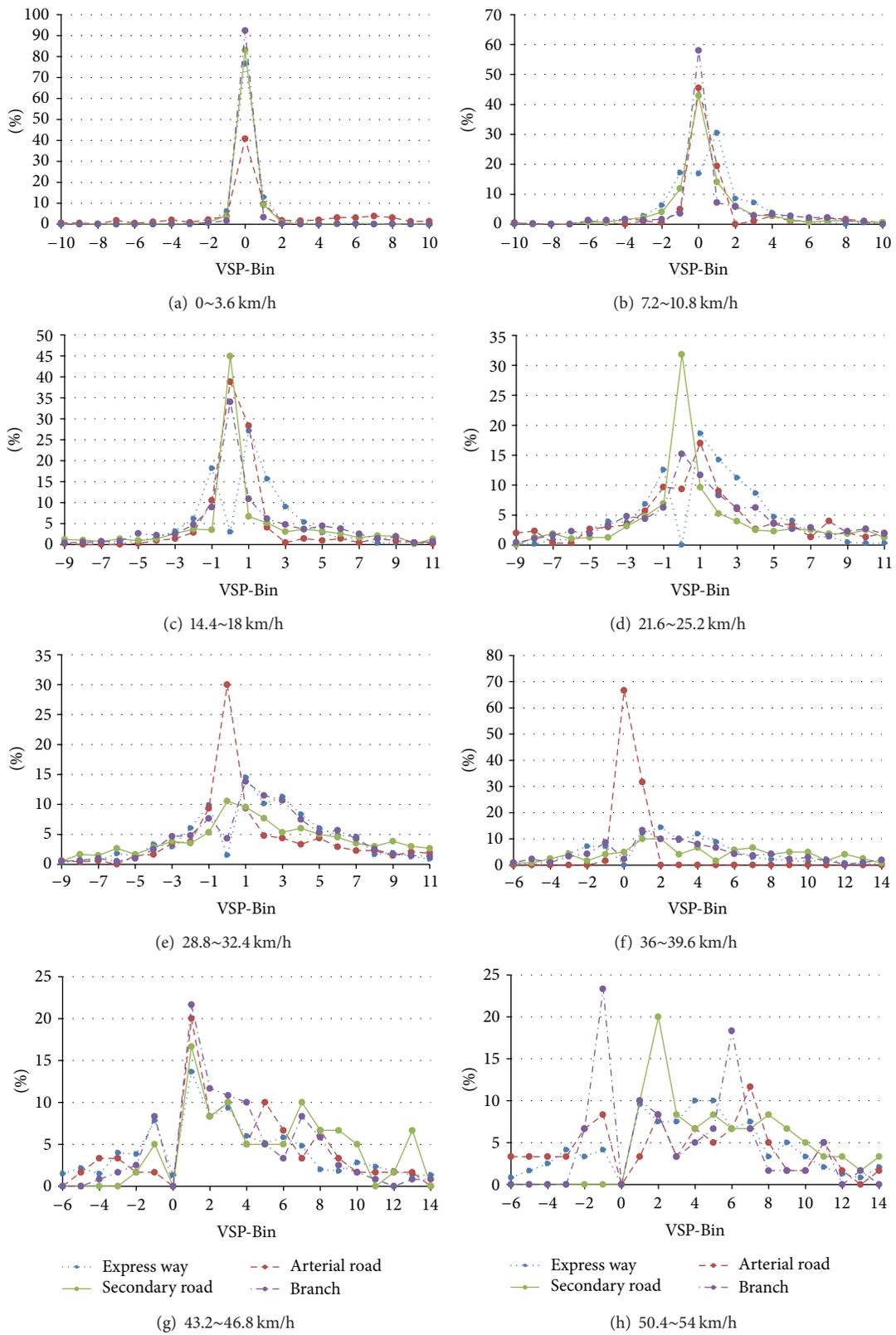


FIGURE 5: VSP-Bin distributions for different road types.

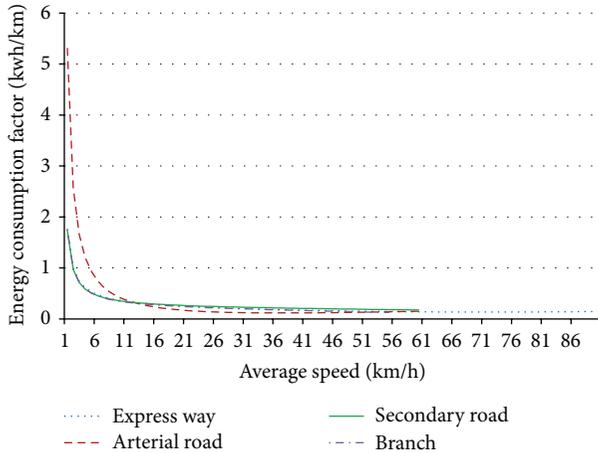


FIGURE 6: Energy consumption factors for different road types.

## 4. Conclusions

This paper presents a methodology for establishing the EV's energy consumption models for different road types. The proposed models are developed through considering the influence of different VSP distribution of different road types on energy consumption, which not only guarantees the accuracy of energy consumption factor models but also makes it possible to estimate the energy consumption for different road types based on current traffic information systems. It can be concluded from the energy consumption factor curves for different road types (shown in Figure 6) that the energy consumption factor of arterial road is notably different from other road types. It is validated that road type-based energy consumption factor should be used in link energy cost calculation in EV's energy-optimal route planning process. In the future practical application, the models presented in this paper can be utilized to forecast EV's cruising range for different road types based on dynamic average travel speed with high accuracy, which can find out the energy-optimal route and reduce the risk of running out of electricity before arriving at the destination or charging station.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research is supported by Research Fund for the Doctoral Program of Higher Education of China (no. 20130009110002) and National Natural Science Foundation of China (no. 71210001).

## References

- [1] C. Musardo, G. Rizzoni, Y. Guezennec, and B. Staccia, "A-ECMS: an adaptive algorithm for hybrid electric vehicle energy management," *European Journal of Control*, vol. 11, no. 4-5, pp. 509–524, 2005.

- [2] T. Zuo, E.-J. Yao, Y. Yang et al., "Study on route planning algorithm for electric vehicle," *Journal of Beijing Institute of Technology*, vol. 21, no. 2, pp. 31–35, 2013.
- [3] M. Barth, K. Boriboonsomsin, and A. Vu, "Environmentally-friendly navigation," in *Proceedings of the 10th International IEEE Conference on Intelligent Transportation Systems (ITSC '07)*, pp. 684–689, Seattle, Wash, USA, October 2007.
- [4] K. Boriboonsomsin and M. Barth, "ECO-routing navigation system based on multi-source historical and real-time traffic information," in *Proceedings of the IEEE Workshop on Emergent Cooperative Technologies in Intelligent Transportation Systems (ICTSC '10)*, 2010.
- [5] H. Wang, L. Fu, Y. Zhou, and H. Li, "Modelling of the fuel consumption for passenger cars regarding driving characteristics," *Transportation Research D*, vol. 13, no. 7, pp. 479–482, 2008.
- [6] A. G. Simpson, *Parametric Modelling of Energy Consumption in Road Vehicles*, University of Queensland, St Lucia, Australia, 2005.
- [7] D. Brzezinski, C. Hart, and P. Enns, "Final facility-specific speed correction factors," Tech. Rep. EPA-420-R-01-060, U.S. Environmental Protection Agency, Washington, DC, USA, 2001.
- [8] Z.-P. Wang, L.-M. Yao, and F.-C. Sun, "Preliminary study on evaluation system of EV energy consumption economy," *Transaction of Beijing Institute of Technology*, vol. 25, no. 6, pp. 479–486, 2005.
- [9] R. Joumard, P. Jost, J. Hickman, and D. Hassel, "Hot passenger car emissions modelling as a function of instantaneous speed and acceleration," *Science of the Total Environment*, vol. 169, pp. 167–174, 1995.
- [10] M. André and C. Pronello, "Relative influence of acceleration and speed on emissions under actual driving conditions," *International Journal of Vehicle Design*, vol. 18, no. 3-4, pp. 340–353, 1997.
- [11] Y.-G. Qi, H.-H. Teng, and L. Yu, "Microscale emission models incorporating acceleration and deceleration," *Journal of Transportation Engineering*, vol. 130, no. 3, pp. 348–359, 2004.
- [12] E.-J. Yao, Z.-F. Lang, Y.-Y. Song, Y. Yang, and T. Zuo, "Microscopic driving parameters-based energy-saving effect analysis under different electric vehicle penetration," *Advances in Mechanical Engineering*, vol. 2013, Article ID 435721, 8 pages, 2013.
- [13] J. Q. Hansen, M. Winther, and S. C. Sorenson, "The influence of driving patterns on petrol passenger car emission," *Science of the Total Environment*, vol. 169, pp. 129–139, 1995.
- [14] T. Zachariadis and Z. Samaras, "Comparative assessment of European tools to estimate traffic emissions," *International Journal of Vehicle Design*, vol. 18, no. 3-4, pp. 312–325, 1997.
- [15] K. Ahn, H. Rakha, A. Trani, and M. Van Aerde, "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels," *Journal of Transportation Engineering*, vol. 128, no. 2, pp. 182–190, 2002.
- [16] J. L. Jiménez-Palacios, *Understanding and quantifying motor vehicle emissions with vehicle specific power and TILDS remote sensing [Ph.D. thesis]*, Massachusetts Institute of Technology (MIT), Cambridge, Mass, USA, 1999.
- [17] T. Younglove, G. Scora, and M. Barth, "Designing on-road vehicle test programs for the development of effective vehicle emission models," *Transportation Research Record*, no. 1941, pp. 51–59, 2005.

- [18] E. K. Nam and R. Giannelli, "Fuel consumption modeling of conventional and advanced technology vehicles in the physical emission rate estimator (PERE)," Tech. Rep. EPA420-P-05-001, Assessment and Standards Division Office of Transportation and Air Quality US Environmental Protection Agency, Washington, DC, USA, 2005.
- [19] G.-H. Song and L. Yu, "Estimation of fuel efficiency of road traffic by characterization of vehicle-specific power and speed based on floating car data," *Transportation Research Record*, no. 2139, pp. 11–20, 2009.
- [20] E.-J. Yao and Y.-Y. Song, "Study on eco-route planning algorithm and environmental impact assessment," *Journal of Intelligent Transportation Systems*, vol. 17, no. 1, pp. 42–53, 2013.
- [21] J. Liu and L. Yu, *Speed correction model for vehicle emissions and fuel consumption based on VSP distributions [M.S. thesis]*, Beijing Jiaotong University, Beijing, China, 2010.
- [22] G.-H. Song, L. Yu, and Z. Tu, "Distribution characteristics of vehicle-specific power on urban restricted-access roadways," *Journal of Transportation Engineering*, vol. 138, no. 2, pp. 202–209, 2011.
- [23] Y. Namikawa, Y. Takai, and N. Ohshiro, *Calculation Base of Motor Vehicle Emission Factors*, Technical Note of National Institute for Land and Infrastructure Management, 2003.

## Research Article

# Efficient Processing of Continuous Skyline Query over Smarter Traffic Data Stream for Cloud Computing

Wang Hanning,<sup>1</sup> Xu Weixiang,<sup>2</sup> Jiulin Yang,<sup>3</sup> Lili Wei,<sup>4</sup> and Jia Chaolong<sup>1</sup>

<sup>1</sup> State Key Laboratory of Rail Traffic Control and Safety, Beijing JiaoTong University, Beijing 100044, China

<sup>2</sup> School of Traffic and Transportation, Beijing JiaoTong University, Beijing 100044, China

<sup>3</sup> China National Tendering Center of Mach. & Elec. Equipment, Beijing 100142, China

<sup>4</sup> Chongqing Public Security Bureau, Chongqing 401147, China

Correspondence should be addressed to Xu Weixiang; [wxxu@bjtu.edu.cn](mailto:wxxu@bjtu.edu.cn)

Received 19 September 2013; Accepted 4 November 2013

Academic Editor: Huimin Niu

Copyright © 2013 Wang Hanning et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The analyzing and processing of multisource real-time transportation data stream lay a foundation for the smart transportation's sensibility, interconnection, integration, and real-time decision making. Strong computing ability and valid mass data management mode provided by the cloud computing, is feasible for handling *Skyline* continuous query in the mass distributed uncertain transportation data stream. In this paper, we gave architecture of layered smart transportation about data processing, and we formalized the description about continuous query over smart transportation data *Skyline*. Besides, we proposed *mMR-SUDS* algorithm (*Skyline* query algorithm of uncertain transportation stream data based on *micro-batchinMap Reduce*) based on sliding window division and architecture.

## 1. Introduction

Recently, tremendous changes have taken place in city transportation data sources, transportation data services, and information infrastructure. Traditional *ITS* (*intelligent transport systems*) present many defects in higher-dimensional space-time continuous data stream collected and passed back from mass perceptible and measurable sensor networks and the storage, processing, and analysis of big data. With the advent of computing technology such as Internet of things, cloud computing [1], and smarter transportation [2] has emerged, as a new concept of comprehensive transportation system. As shown in Figure 1, smarter transportation system covers various aspects of transportation and is a complex and comprehensive system consisting of plenty of subsystems. Analytical processing of multi-source and real-time transportation data stream [3] is the basis of realizing perceptible Smarter Transportation with interconnection integration and real-time decision. Besides, such analytical processing is critical to establishing global sustainable transportation surveillance, network optimization

of dynamic transportation, automatic response to accidents, and integration of location-based transportation services.

With the rapid development of information technology, monitoring platform in various types of transportation information management collects complex mass transportation stream data including video information [4, 5] from cameras, monitoring information of sensors, positioning system information of vehicle, and so on. Hence, transportation stream data are provided with diverse sources, wide varieties, different forms, and typical data-intensive processing characteristics. For example, by December 28, 2012, there were 8842 fixed transportation monitoring equipment in Beijing and merely dispatch center for transportation operational monitoring *TOCC* in Beijing updated over 3500 data immediately and replaced more than 20 thousand video pictures in real time. Operational applications of environmental sensor station are shown in Figure 2. Real-time transportation data stream lays important data foundation for road transportation stream control of various decision analysis and emergency response in smarter transportation system. *Skyline* [6] query, as a key data mining technology,

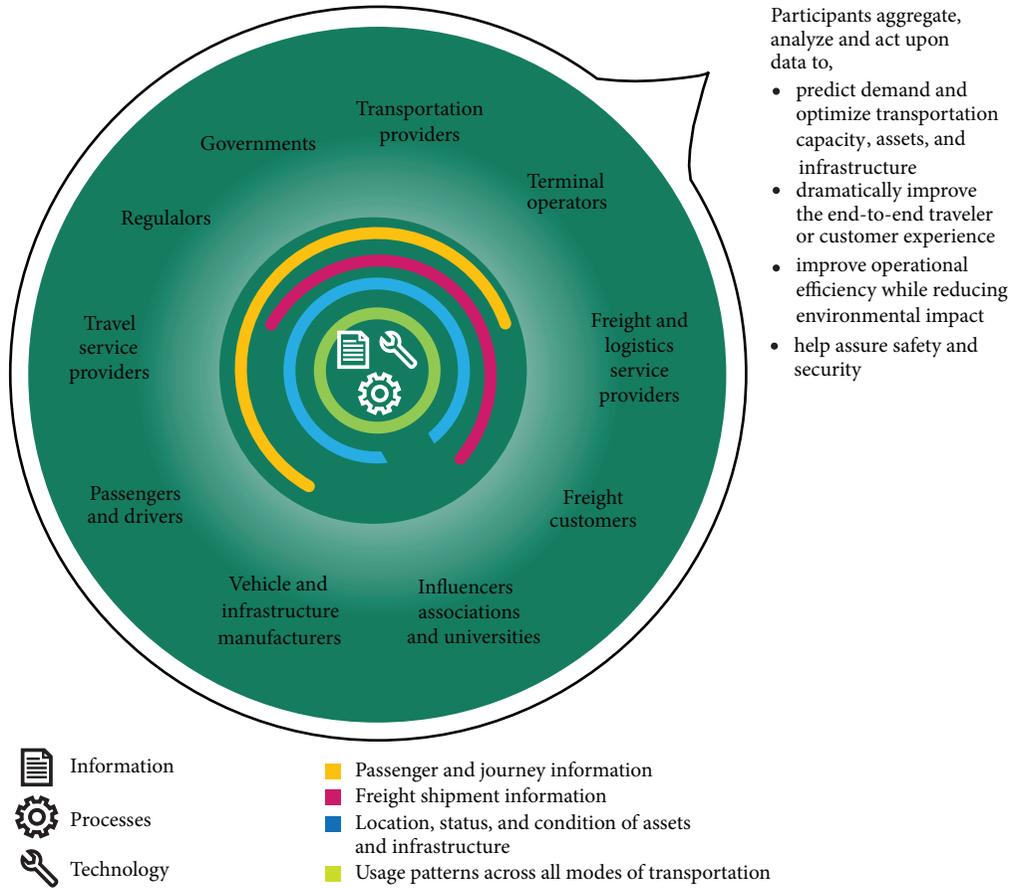


FIGURE 1: Smarter Transportation Information.

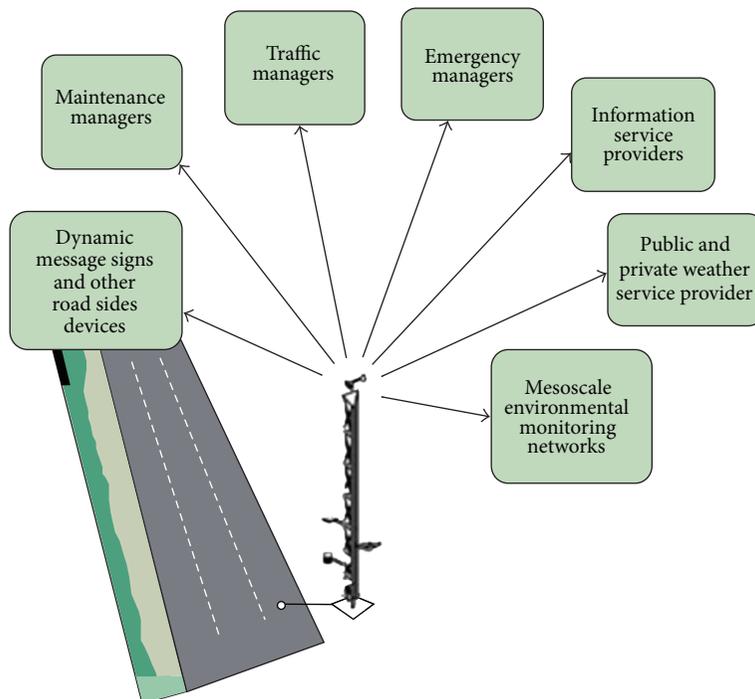


FIGURE 2: Operational application of environmental sensor station.

is of great significance in multiconstrained decision support, city navigation, user preference query, visualization of data mining, and so on under dynamic environment [7–9]. Hence, such query is consistent with practical application of data stream processing of smarter transportation. In addition, collection and analytical processing of transportation stream data present geographically distributed characteristic and are often influenced by uncertain sources such as wireless sensor networks, wireless radio frequency identification, location-based services, moving object management, and so on. Thus, data objects in data stream present uncertainty. Therefore, uncertain [10] real-time transportation data stream is characterized by difficult prediction, variability, rapid arrival, mass and infinite arrival, and so forth. Meanwhile, analytical processing of transportation stream data requires multiservice parallel processing and very high timeliness. In the environment of cloud computing, this paper combined the processing requirements for complex, parallel, and real-time transportation stream data and investigated continuous *Skyline* query algorithm with low cost, rapid response, and efficient scalability based on parallel processing framework of mass data. Compared with traditional *Skyline* query, *Skyline* query over uncertain transportation stream data faces the following challenges.

- (1) In computational process of *Skyline* query on uncertain data stream, both dominant relations between computing objects and *Skyline* probability need to be calculated. However, traditional strategies fail to perform this process directly. Obviously, *Skyline* query calculation is CPU intensive [11, 12] and very high processing ability is required.
- (2) Transportation stream data arrives continuously and is required to be processed immediately. So, when data stream is too rapid and users pay attention to a great number of objects (sliding window [13] is very large), traditional algorithm of centralized stream processing is difficult to satisfy the query demand.

Cloud computing with high storage capacity and calculating ability can fully satisfy application requirements of *Skyline* query on mass data. Main contributions of this research are as follows.

- (1) Processing architecture of stratified transportation stream data is demonstrated.
- (2) In the environment of cloud computing, this paper proposes the issue of continuous *Skyline* query on mass distributed uncertain transportation data stream and provides formal description.
- (3) This research develops an *mMR-SUDS* algorithm based on sliding window division and the architecture proposed.

Section 2 introduces the processing architecture of stratified stream data of smarter transportation, background information, relevant work, and formal description of the problem. Section 3 explains design conception and optimization strategy of *mMR-SUDS* algorithm. Besides, experimental result comparison is demonstrated in Section 4, while summary of the entire research is made in Section 5.

## 2. Setting

**2.1. Processing Architecture of Stratified Transportation Stream Data.** In smarter transportation, processing architecture of stratified transportation stream data is shown in Figure 3. Bottom layer is front end of perceptible equipment consisting of  $N$  acquisition nodes for remote real-time data monitoring. Interlayer consists of  $M$  coordinator nodes connected to high speed network, while all transportation data processing centers are placed on top layer, providing transportation data services such as control, analysis, early warning, and so on.

**2.2. Relevant Work.** Early *Skyline* query is commonly applied to centralized database. Relevant researches mainly focus on centralized algorithms such as *block-nested-loops*, *BNL* algorithm [6]; *divide-and-conquer*, *D&C* algorithm [6]; *sort-filter-Skyline*, *SFS* algorithm [14]; *nearest neighbor*, *NN* algorithm [15]; *branch-and-bound Skyline*, *BBS* algorithm [16]; *bitmap* algorithm [17], and so on. Jian et al. [18] first proposed *Skyline* query technology on uncertain data and presented two query algorithms: bottom-up algorithm and top-down algorithm. In addition, in terms of uncertain data presentation, relevant researches usually pay more attention to discrete data. Therefore, according to literature [19], based on uncertain data at attribute level, three defined constraint methods including *uncertainty reduction*, *pairwise comparison*, and *adaptive bound tightening* were proposed to optimize *Skyline* query calculation.

In the field of *Skyline* query over data stream, aimed at continuous *Skyline* query based on sliding window model, literature [20] proposed *Lazy* algorithm and *Eager* algorithm which improves space and time efficiency using the method of advanced data cleaning. In addition, literature [21] investigated *Skyline* query of *n-of-N* data stream model in sliding window and proposed continuous *n-of-N* algorithm that improves system space performance by defining “key domination.”

In the field of *Skyline* query over uncertain data stream, the data model in literature [22] was a data set consisting of certain objects where variable amounts of examples were presented for each object. And the concept of *Skyline* probability was proposed based on *Skyline* probability of examples for each object. Hence, the data model in literature [23] was virtually a discretionary version of uncertain attribute, while this paper focused on the case of uncertain tuple. On the other hand, literature [24] concentrated on static dataset, while this paper concentrated on data stream. Moreover, aimed at efficient *Skyline* calculation of uncertain data stream, literature [13] proposed *Skyline* query based on probability threshold and used the optimization methods like *Skyline* candidate sets and so on to execute continuous *Skyline* query efficiently. In contrast, literature [25] presented *Skyline* over probabilistic data stream algorithm. Based on grid index with better adaptability, heuristic rules such as probability delimitation, stepwise refinement, elimination in advance, optional indemnity, and so on were employed to optimize the algorithm temporally and spatially. By comparison, literature [26] investigated expectation evaluation of *Skyline* probability

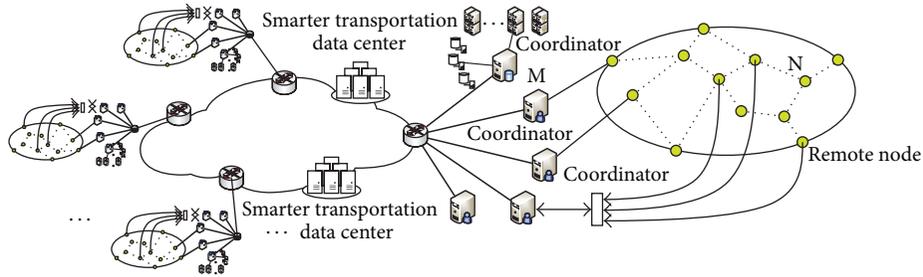


FIGURE 3: Processing architecture of stratified transportation stream data.

and presented the relation between probability threshold and expectation of *Skyline* probability.

In the field of distributed parallel *Skyline* query, current researches mainly focused on static data. Literature [27] suggested that integral query performance of system could be improved by defining execution order of *Skyline* query on each server. In addition, parallel distributed *Skyline* algorithm proposed in literature [28, 29] divided relevant sites into several groups by data division method and queries among groups were executed in parallel.

According to processing requirements of mass data, several existing relevant researches combined *Map Reduce* technology with *Skyline* query algorithm. Literature [4] proposed preview *Skyline* query algorithm and attempted to reduce size of input data in *Map* task and *Reduce* task through preview filtration. Thus, the performance of *Skyline* query based on *Map Reduce* framework was improved.

### 2.3. Terms and Definition

**2.3.1. Data Stream.** In a formal way, a data stream is any ordered pair  $(s, \Delta)$  where  $s$  is a sequence of tuples and  $\Delta$  is a sequence of positive real time intervals. For instance, there is a data stream with following tuple model in management system of road transportation stream (see Figure 4).

*Road Stream* is defined as data stream of tuple model processed in data stream management system. In the tuple model, attribute *Road Stream* denotes the name of the data stream, while *Vehicle\_ID* denotes the unique identifier of a vehicle. Moreover, *X\_Way* denotes road section of a vehicle; *X\_Pos* presents the location of a vehicle; *Express\_Way* denotes the expressway number; *Speed* denotes the current speed of a vehicle; *Timestamp* denotes that, when relevant information dispatched by a vehicle arrives at data stream system, system assigns a value to  $T$  according to the time sequence of received information.

### 2.3.2. Skyline

**Definition 1.** *Skyline* A point  $p \in S$  is said to dominate another point  $q \in S$ , denoted as  $p < q$ , if (1) in every dimension  $d_i \in D$ ,  $p_i \leq q_i$ ; (2) in at least one dimension  $d_j \in D$ ,  $p_j < q_j$ . The *Skyline* is a set of points  $SKY(S) \subseteq S$  which are not dominated

by any other point. The points in  $SKY(S)$  are called *Skyline* points.

**Definition 2.** The *Skyline* probability of an instance  $p$ , that is,  $\Pr_{sky}(p)$ , is the probability that  $p$  exists and no instance of other uncertain objects that dominates  $p$  exists. Let  $m$  be the total number of uncertain objects and let  $p \in O_k$ ; we have

$$\Pr_{sky}(p) = \Pr(p) \cdot \prod_{i=1, i \neq k}^m \left( 1 - \sum_{q \in O_i, q < p} \Pr(q) \right). \quad (1)$$

**Definition 3.** Given a dataset  $S$  with  $n$  instances that belong to  $m$  uncertain objects and a probability threshold  $\vartheta$ , the instance-level probabilistic *Skyline* analysis returns all instances with *Skyline* probabilities at least  $\vartheta$ . That is, return the *Skyline* set  $S_{sky}$  such that

$$S_{sky} = \{p \in S \mid \Pr_{sky}(p) \geq \vartheta\}. \quad (2)$$

## 3. Skyline Query Algorithm (*mMR-SUDS*) of Uncertain Transportation Stream Data Based on *micro-batchinMap Reduce* Framework

**3.1. Division of Sliding Window.** According to the architecture of distributed transportation stream data processing, coordinator nodes collect continuous uncertain data stream monitored by each remote monitoring node. In this paper a cross method using count sliding window model divided whole sliding window so that data in the whole large sliding window of uncertain data stream are divided effectively. Then, data were distributed to various parallel computational nodes in order that each parallel computational node could actually correspond to a valid part of the whole sliding window. The basic conception was as follows: coordinator nodes dispatch arrived data successively to parallel nodes, and each parallel node maintains a count sliding window part. Thereby, the sliding window parts on all parallel nodes are combined across in turn, logically corresponding to the whole sliding window of uncertain data stream. And the corresponding relations are shown in Figures 5 and 6.

Roadstream	Vehicle_ID	X_pos	X_way	Express_way	Dir	Speed	Timestamp
------------	------------	-------	-------	-------------	-----	-------	-----------

FIGURE 4

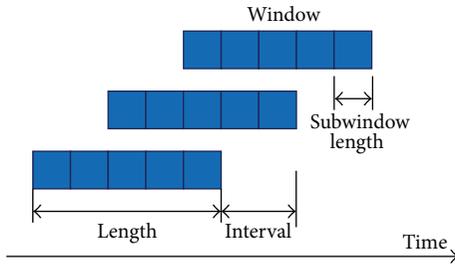


FIGURE 5: Sliding windows and Subwindows.

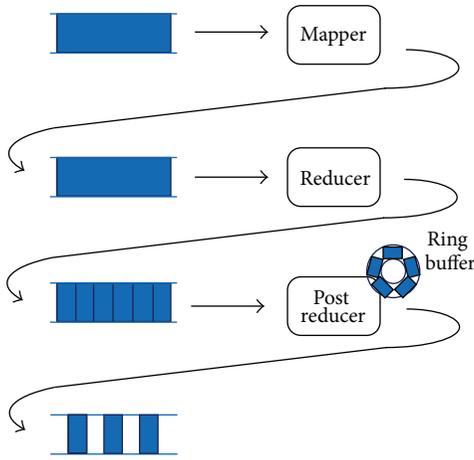


FIGURE 6: Sliding windows Implementation.

### 3.2. Processing Framework of Transportation Stream Data.

Based on the sliding window division and *micro-batchinMap Reduce* model, processing framework of transportation stream data is designed in Figure 7. The framework consists of four types of nodes: *Coordinator* nodes that are responsible for reception of input data stream and data dispatch to *Map-PE* nodes (map-processing element); *Map-PE* nodes that are responsible for maintenance of data refresh in sliding window of *Map-PE* nodes and calculation of *Skyline* probability presented in the form of  $MP_1, MP_2 \dots MP_n$ , which can mutually communicate with each other; *Reduce-Q* nodes (*reduce-query*) that are responsible for reception of *Skyline* results from each computational node; and Master nodes that are responsible for status maintenance of *Map-PE* nodes and *Reduce-Q* nodes. Besides,  $u$ ,  $v$ , and  $w$  denote investigated uncertain data. According to processing framework of parallel data stream based on division of sliding window, *Skyline* query process of uncertain smarter transportation stream data is as follows.

- (1) When uncertain data  $u$  arrives at *Coordinator* nodes, *Coordinator* nodes dispatch  $u$  to *Map-PE* node  $MP_1$ .

- (2)  $MP_1$  maintains renewed variation of *Skyline* probability caused by overdue data  $v$  and incoming data  $u$  in the window of *Map-PE* node. Then,  $MP_1$  node dispatches overdue data  $v$  and newly incoming data  $u$  to other *Map-PE* nodes.
- (3) Each *Map-PE* node maintains renewed variation of *Skyline* probability resulting from overdue data  $v$  and newly incoming data  $u$  in the window of each *Map-PE* node. This type of nodes is only in charge of updating *Skyline* probability and sending the updated results to *Reduce-Q* nodes. And all the parallel nodes dispatch feedback about *Skyline* probability of data  $u$  in the corresponding node to  $MP_1$ .
- (4) Taking the feedbacks from all nodes about *Skyline* probability of data  $u$  into account,  $MP_1$  calculates global *Skyline* probability of data  $u$  and outputs the result to query nodes.
- (5) When new uncertain data  $w$  arrives at *Map-PE* nodes, *Map-PE* nodes dispatch  $w$  to  $MP_2$  which performs the above mentioned process circularly.

**3.3. *mMR-SUDS* Algorithm.** The basic conception of *Skyline* query algorithm on uncertain transportation stream data based on *micro-batchinMap Reduce* framework is as follows. The task of updating *Skyline* probability of uncertain transportation data tuple in the whole sliding window is distributed to each parallel node. Then, parallelism among *Map-PE* nodes is employed to improve the operational efficiency of overall system. Hence, algorithm realization of all types of nodes is discussed in this section.

*Coordinator* nodes are responsible for data cache and data dispatch. Processing algorithm on *Coordinator* nodes is illustrated as follows.

**Input.** Uncertain data stream; response message of all parallel nodes,

**Output.** Data block of uncertain data.

- (1) *Coordinator* nodes receive and then cache the incoming uncertain transportation stream data.
- (2) If *Coordinator* nodes receive response message from a *Map-PE* node, the following results will be presented.
  - (2.1) New data block is obtained from the cache.
  - (2.2) Data are dispatched to the next *Map-PE* node.

Data cache and data dispatch are two procedures executed in parallel in the algorithm above mentioned. Communication of *Coordinator* nodes is followed by the corresponding *Reduce-Q* nodes. *Reduce-Q* nodes are in charge of receiving, synchronizing, and then displaying *Skyline* results dispatched

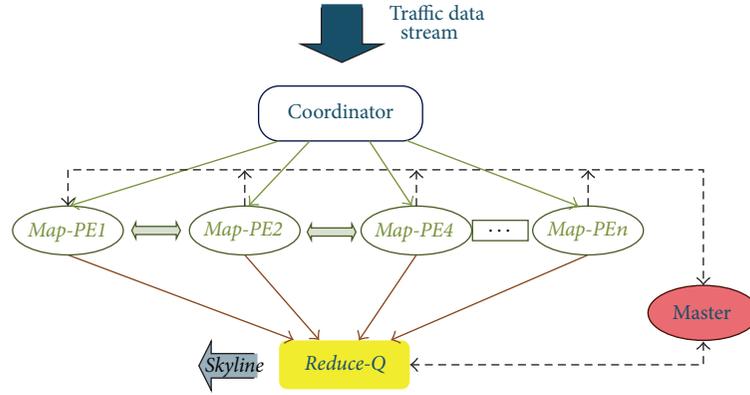


FIGURE 7: Processing framework of transportation stream data.

from all parallel *Map-PE* nodes. Processing algorithm on *Reduce-Q* nodes is as follows.

*Input.* *Skyline* results dispatched from all *Map-PE* nodes.

*Output.* Global *Skyline* results.

- (1) *Skyline* results from all *Map-PE* nodes are received and cached.
- (2) Received information is synchronized and global *Skyline* results are output.

Finally, taking  $MP_i$  as an example, processing algorithm on parallel *Map-PE* nodes is presented as follows.

*Input.* Data block of uncertain data; feedbacks from *Map-PE* nodes.

*Output.* Local *Skyline*; global *Skyline*.

- (1)  $MP_i$  receives and analyzes information.
- (2) If newly incoming data from *Coordinator* nodes are received, the results are as follows.
  - (2.1) New data tuple is obtained from information.
  - (2.2) Overdue tuple is obtained from the current window.
  - (2.3) *Skyline* probability variation caused by overdue data is updated.
  - (2.4) *Skyline* probability variation caused by newly incoming data is updated.
  - (2.5) Local *Skyline* probability of newly incoming data is calculated.
  - (2.6) *Skyline* probability variation in data block caused by dominance relation is calculated.
  - (2.7) Data block is added to local window.
  - (2.8) Updated information including newly incoming tuple and overdue tuple is dispatched to other *Map-PE* nodes.
- (3) Otherwise, if updated information from a *Map-PE* node is received, the following results are presented.

- (3.1) New data tuple is obtained from information.
- (3.2) Overdue data tuple is obtained from information.
- (3.3) *Skyline* probability variation caused by overdue data is updated.
- (3.4) *Skyline* probability variation caused by the arrival of new data is updated.
- (3.5) Local *Skyline* probability of new data is calculated.
- (3.6) Feedbacks including *Skyline* probability of newly incoming tuple in this node are dispatched to nodes transmitting the updated information.
- (3.7) Local *Skyline* results are dispatched to *Reduce-Q* nodes.

- (4) Otherwise, if feedbacks from a *Map-PE* node are received, consolidated calculation is performed.

- (4.1) New data tuple is obtained from information and local *Skyline* probability of new tuple is calculated.
- (4.2) *Skyline* probability is updated.
- (4.3) If feedbacks from all the *Map-PE* nodes are collected, the results are as follows.
  - (4.3.1) *Skyline* results are dispatched to *Reduce-Q* nodes.

- (5) Otherwise, if node  $i$  receives unrecognized command, error message is presented.

### 3.4. The Optimization of Algorithm

(1) *Reduction of Window Scanning Times.* When analyzing the processing algorithm on parallel computational nodes, it can be found that three times of window scanning were presented, respectively, in procedures 2.3, 2.4, and 2.5. Besides, there were also three times of window scanning, respectively, in procedures 3.3, 3.4, and 3.5. To reduce window scanning times, three times of window scanning can be integrated into one time scanning. Moreover, in each window scanning, data in the window is compared with new data and overdue data.

Thus, processing performance of the algorithm is improved by reducing window scanning times.

(2) *Intermediate Filtration*. Computational process of data *Skyline* probability shows

$$P_{sky}(a) = P(a) \times P_{old}(a) \times P_{new}(a). \quad (3)$$

That is, *Skyline* probability of tuple  $a$   $P_{sky}(a)$  equals the product of three probabilities including existing probability of tuple  $a$   $P(a)$ , probability of tuple  $a$  not dominated by the data arriving earlier  $P_{old}(a)$  and probability of tuple  $a$  not dominated by the data arriving later  $P_{new}(a)$ . Among the three probabilities, with new data arriving and old data expiring,  $P_{old}(a)$  increases continuously, while  $P_{new}(a)$  decreases constantly. Moreover,  $P(a)$ ,  $P_{old}(a)$ , and  $P_{new}(a)$  are all in the interval  $(0, 1)$  throughout. Therefore, if  $P_{new}(a) < \rho$ , the relation that  $P_{sky}(a) < \rho$  is established. In addition, during the life cycle of  $a$  (time when  $a$  is in the sliding window),  $P_{sky}(a) < \rho$  is established permanently so it is unnecessary to calculate  $P_{sky}(a)$ . Hence, through the method of intermediate filtration, times of comparison are reduced and algorithm processing speed is accelerated owing to the fact that result set is far less than source dataset.

(3) *Decrease of Idle Waiting Time of Nodes*. It is presumed that all *Map-PE* nodes are provided with the same processing ability.  $t$  denotes average time that a node takes to communicate once with another node, while  $T$  denotes average time of one calculation update of *Skyline* probability except consolidated calculation. Besides,  $T'$  denotes average time of consolidated calculation. And the relation of the three is that  $t < T < T'$ . In basic scheme, calculation period of *Skyline* probability update caused by one data update is shown in Figure 8.

Figure 8 indicates that, when  $MP_1$  receives newly incoming data, local *Skyline* probability update is achieved first and then updated information is dispatched to other *Map-PE* nodes. Therefore,  $MP_1$  is completely in idle waiting state before consolidated calculation and idle waiting time of  $MP_1$  is  $T + t$ . Similarly, it can be obtained that idle waiting time of other nodes is  $T' + T + 2t$ . So, in this condition, it takes  $T' + 2T + 3t$  to complete an entire calculation period.

To decrease idle waiting time of all *Map-PE* nodes, when receiving newly incoming data, *Map-PE* nodes can dispatch updated information to other parallel nodes first and then calculate local *Skyline* probability for update. The revised calculation period is illustrated in Figure 9.

Figure 9 shows that, in optimized scheme, idle waiting time of  $MP_1$  is  $t$  and that of other parallel nodes is  $T' + 2t$ . As a result, it takes  $T' + 2T + 3t$  to achieve a complete calculation period. Therefore, compared with basic scheme, optimized scheme saves  $T$  in a calculation period.

## 4. Experimental Evaluation

Algorithm in this paper was realized using Java language and experiments were conducted in practical data-centered environment. Every processing node was configured with a CPU of Pentium4 with 2.0 GHz, a DDR memory of 2 GB,

and Ubuntu operating system. Besides, synthetic data (characterized as independently distributed data) in literatures was adopted in experimental tests and existing probability of tuples followed Gaussian distribution. In synthetic data, data in all dimensions is mutually independent and presents uniform distribution in the interval  $[0, 1]$ . To test the real processing performance of *mMR-SUDS* algorithm, this paper presumed that *Coordinator* nodes cache numerous data tuples. When parallel nodes finish processing a batch of data and dispatch data request to *Coordinator* nodes, *Coordinator* nodes dispatch new stream data to parallel nodes for processing. In addition, probability threshold in experiments was set to 0.3 and window length was measured by data tuples contained in window with the value of 10000, 100000, 500000, and 1000000, respectively. Ranges of other experimental parameters were as follows: data dimension ranged from 2 to 6; the size of transmission data block (the number of data) was set to 1, 10, 100, and 1000, respectively, while the number of nodes participating in calculation was set to 1, 2, 4, 8, and 16, respectively. Each group of experiments was conducted 10 times and the average value was taken as the result. In contrast experiments, as a single machine algorithm, *Base* algorithm includes two nodes: a data cache node and a computational node. The data cache node is responsible for data cache and data dispatch to the computational node, while the computational node is responsible for the maintenance of sliding window and *Skyline* calculation. Besides, the computational node adopts the method of circularly dominating comparison. That is, once data arrives or expires, the computational node compares the incoming data or overdue data with all the data in sliding window and then updates *Skyline* probability.

Based on the experimental environment and experimental data above mentioned, the performance of *mMR-SUDS* algorithm was tested, respectively, in different sizes of transmission data block, window length, data dimension, and number of nodes.

*4.1. Influence Tests of Transmission Data Block*. Uncertain data stream is transmitted in data block between *Coordinator* nodes and *Map-PE* nodes as well as between *Map-PE* nodes. Therefore, size of transmission data block has a certain influence on algorithm realization. To evaluate such influence, this group of experiments tested the algorithm performance in different sizes of transmission block. In experiments, transmission data block was set to 1, 10, 100, and 1000, respectively; data dimension was set to 2, while window length was set to 1000000. Moreover, there were 16 *Map-PE* parallel nodes participating in the calculation.

Experimental results are demonstrated in Figure 10. With the constant increase of transmission data block, processing speed of *mMR-SUDS* algorithm tends to increase first and then decrease. The main reasons are as follows: when data block is small, overhead communication increases due to frequent data transmission, while when transmission data block is large, computation cost increases due to the increasing complexity of data dominated comparison in block. In conclusion, when size of transmission data block takes the

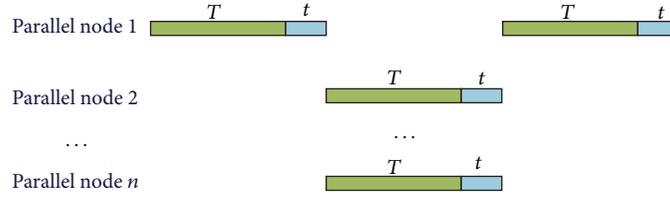


FIGURE 8: The computing cycle in the basic scheme.

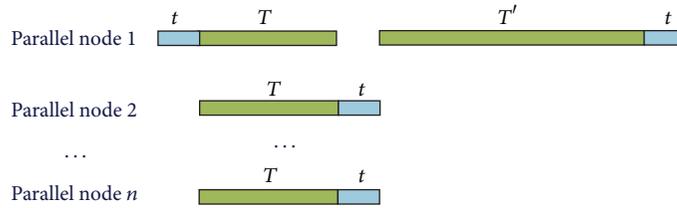


FIGURE 9: The computing cycle in the improved scheme.

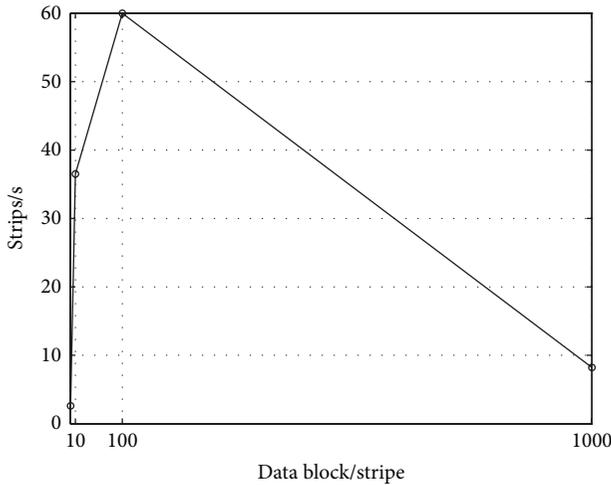


FIGURE 10: The effect of the size of transmitted data block.

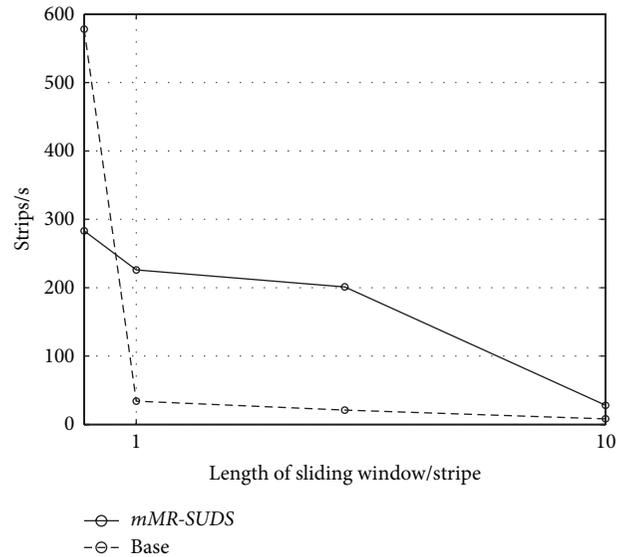


FIGURE 11: The effect of the length of sliding window.

middle value of 100, the algorithm provides good processing performance.

**4.2. Tests of Window Scalability.** In this group of experiments, data dimension was set to 2 and size of transmission data block was set to 100, while window length ranged from 10000 to 1000000. To compare the performance of *mMR-SUDS* algorithm with that of *Base* algorithm, 16 *Map-PE* parallel nodes participated in the calculation.

Experimental results are illustrated in Figure 11. As window length increases constantly, system processing performance declines rapidly. When window length is 10000, performance of *Base* algorithm is even better than that of *mMR-SUDS* algorithm. The main reasons are as follows. When window length is small, calculating performance of single machine fully satisfies the requirement of query processing. But in parallel algorithm, in terms of the whole parallel computing system, much time is taken to deal with

problems such as communication, synchronization, and so on, although each node participating in calculation completes query processing rapidly. When window length is 100000 or more, single computational node could not fully satisfy the performance requirement of query processing. And for the whole parallel computing system, time overhead is mainly spent on calculation and parallel computing system begins to present the advantage of parallelism.

**4.3. Tests of Dimension Scalability.** To compare the dimension scalability of *mMR-SUDS* algorithm with that of *Base* algorithm, window length was set to 1000000 and there were 16 parallel processing nodes in system. In addition, data dimension value was in the interval [2, 11] and size of transmission data block was set to 100 in this group of experiments. And Figure 12 demonstrates the experimental

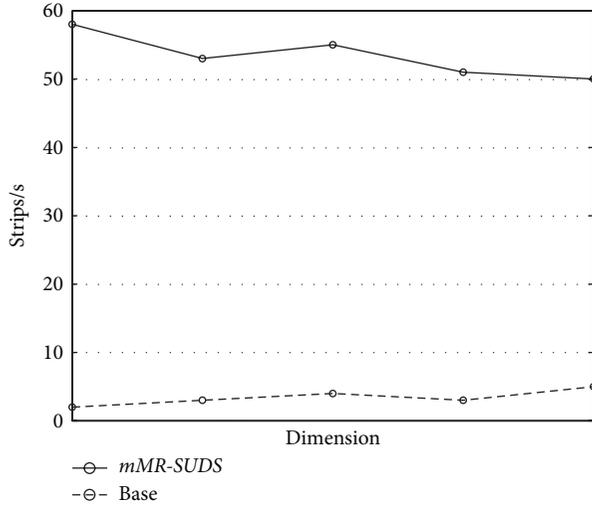


FIGURE 12: The effect of the dimension of data.

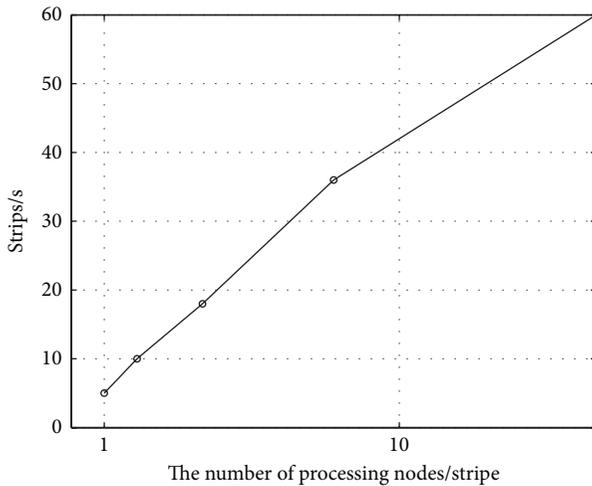


FIGURE 13: The effect of the number of processing nodes.

results. With the increase of data dimension, processing speed of both *mMR-SUDS* algorithm and *Base* algorithm declines slowly; but processing speed of *mMR-SUDS* algorithm is about 12 times higher than that of *Base* algorithm throughout. All in all, *mMR-SUDS* algorithm provides better dimension scalability.

**4.4. Parallel Scalability Tests.** To evaluate the parallel scalability of *mMR-SUDS* algorithm, this group of experiments tested processing performance of the algorithm in different numbers of nodes. In the experiments, the number of parallel nodes took the values of 1, 2, 4, 8, and 16, respectively, and total length of window was set to 1000000. Moreover, data dimension was set to 2, while size of transmission data block was set to 100.

Experimental results are illustrated in Figure 13. As the number of nodes increases continuously, processing speed of *mMR-SUDS* algorithm constantly increases, but the increasing range gradually decreases. The main reasons are as

follows: with the increasing number of nodes, window length on each node decreases gradually. Hence, computation cost of each computational node gradually declines, while overhead communication gradually increases, which influences system processing performance. When the number of nodes took the value of 16, processing ability of *mMR-SUDS* algorithm was about 12 times better than that of single machine algorithm. And processing ability of *mMR-SUDS* in this case was far less than the theoretically optimum value which is as 16 times as that of single machine algorithm. When the number of nodes took the value of 2, processing ability of *mMR-SUDS* algorithm was the closest to the theoretically optimum value that was nearly twice that of single machine algorithm.

## 5. Conclusion

Aimed at *Skyline* query requirements of real-time uncertain data stream of smarter transportation with high capacity and large sliding window in the environment of cloud computing, this paper proposed a *Skyline* query algorithm *mMR-SUDS* over uncertain transportation stream data based on *microbatchinMap Reduce* framework. Such algorithm transforms centralized processing problem of the whole global sliding window into the parallel processing problem of many nodes to their corresponding window by dividing data in sliding window. And such transformation effectively improves integral query processing performance. Experimental results show that *mMR-SUDS* algorithm presents not only high efficiency but, good scalability and load balancing. Therefore, such algorithm could satisfy the processing analysis requirements of various real-time transportation stream data.

In the parallel framework based on sliding window division, future research has to further optimize processing algorithm and improve algorithm processing performance using index structures such as grid, R tree and so on. Meanwhile, research scope of uncertain data shall be expanded to investigate *Skyline* query processing algorithm over uncertain transportation stream data at attribute level.

## Conflict of Interests

The authors declare that they have no financial and personal relationships with other people or organizations that can inappropriately influence their work; there is no professional or other personal interest of any nature or kind in any product, service, and/or company that could be construed as influencing the position presented in, or the review of, the paper.

## Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant no. 61272029), the National Key Technology R&D Program (Grant no. 2009BAG12A10), and independent subject of State Key Laboratory of Rail Traffic Control and Safety, Beijing JiaoTong University (Contract no. RCS2009ZT007) and partially supported by the MOE key Laboratory for Transportation Complex Systems Theory and Technology School of Traffic and Transportation, Beijing, JiaoTong University.

## References

- [1] M. Armbrust, A. Fox, R. Griffith, and M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing," Rep UCB/EECS-2009-28, UC, RAD Laboratory, Berkeley, Calif, USA, 2009.
- [2] IBM, "The Case for Smarter Transportation," Whitepaper, IBM-Institute for Business Value, 2010.
- [3] B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom, "Models and issues in data stream systems," in *Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS '02)*, pp. 1–16, New York, NY, USA, June 2002.
- [4] H. Guo, W. Wang, W. Guo, X. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [5] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: a fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.
- [6] S. Borzsonyi, D. Kossmamm, and K. Stocker, "The skyline operator," in *Proceedings of the 17th International Conference on Data Engineering*, IEEE Computer Society, Washington, DC, USA, 2001.
- [7] C. Jia, W. Xu, F. Wang, and H. Wang, "Track irregularity time series analysis and trend forecasting," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 387857, 15 pages, 2012.
- [8] H. Wang, W. Xu, F. Wang, and C. A. Jia, "cloud-computing-based data placement strategy in high-speed railway," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 396387, 15 pages, 2012.
- [9] J. I. A. Chaolong, X. U. Weixiang, W. E. I. Lili et al., "Study of railway track irregularity standard deviation time series based on data mining and linear model," *Mathematical Problems in Engineering Volume*, vol. 2013, Article ID 486738, 12 pages, 2013.
- [10] C. K. -S. Leung, "Mining uncertain data," *Wiley Interdisciplinary Reviews*, vol. 1, no. 4, pp. 316–329, 2011.
- [11] G. Juve, E. Deelman, K. Vahi et al., "Scientific workflow applications on amazon EC2," in *Proceedings of the 5th IEEE International Conference on e-Science Workshops (e-science '09)*, pp. 59–66, Oxford, UK, December 2009.
- [12] E. Wu, Y. Diao, and S. Rizvi, "High-performance complex event processing over streams," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 407–418, June 2006.
- [13] W. Zhang, X. Lin, Y. Zhang, W. Wang, and J. X. Yu, "Probabilistic skyline operator over sliding windows," in *Proceedings of the 25th IEEE International Conference on Data Engineering (ICDE '09)*, pp. 1060–1071, IEEE Computer Society, Shanghai, China, April 2009.
- [14] J. Chomicki, P. Godfrey, J. Gryz, and D. Liang, "Skyline with presorting," in *Proceedings of the 19th International Conference on Data Engineering*, pp. 717–719, IEEE Computer Society, Bangalore, India, March 2003.
- [15] D. Kossmann, F. Ramsak, and S. Rost, "Shooting stars in the sky: an online algorithm for Skyline queries," in *Proceedings of the 28th International Conference on Very Large Data Bases (VLDB '02)*, pp. 275–286, Hong Kong, 2002.
- [16] D. Papadias, Y. Tao, G. Fu, and B. Seeger, "An optimal and progressive algorithm for skyline queries," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '03)*, pp. 467–478, San Diego, Calif, USA, June 2003.
- [17] K. L. Tan, P. K. Eng, and B. C. Ooi, "Efficient progressive Skyline computation," in *Proceedings of the 27th International Conference on Very Large Data Bases (VLDB '01)*, pp. 301–310, San Francisco, Calif, USA, 2001.
- [18] P. Jian, J. Bin, L. Xuemin et al., "Probabilistic Skylines on uncertain data," in *Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB '07)*, pp. 15–26, Vienna, Austria, 2007.
- [19] M. E. Khalefa, M. F. Mokbel, and J. J. Levandoski, "Skyline query processing for uncertain data," in *Proceedings of the 19th International Conference on Information and Knowledge Management and Co-located Workshops, (CIKM '10)*, pp. 1293–1296, Toronto, Canada, October 2010.
- [20] Y. Tao and D. Papadias, "Maintaining sliding window skylines on data streams," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 3, pp. 377–391, 2006.
- [21] X. Lin, Y. Yuan, W. Wang, and H. Lu, "Stabbing the sky: efficient skyline computation over sliding windows," in *Proceedings of the 21st International Conference on Data Engineering (ICDE '05)*, pp. 502–513, IEEE Computer Society, Tokyo, Japan, April 2005.
- [22] W. Xiaowei, H. Jiuming, and J. Yan, "Probabilistic skyline computation on distributed uncertain data," *Journal of Frontiers of Computer Science and Technology*, vol. 4, no. 10, pp. 951–961, 2010.
- [23] X. Chuanfei, L. Shukuan, W. Lei, and Q. Jianzhong, "Complex event detection in probabilistic stream," in *Proceedings of the 12th International Asia Pacific Web Conference (APWeb '10)*, pp. 361–363, April 2010.
- [24] X. Ding, X. Lian, L. Chen, and H. Jin, "Continuous monitoring of skylines over uncertain data streams," *Information Sciences*, vol. 184, no. 1, pp. 196–214, 2012.
- [25] S.-L. Sun, D.-B. Dai, Z.-H. Huang, Q.-X. Zhang, and L.-X. Zhou, "Algorithm on computing skyline over probabilistic data stream," *Acta Electronica Sinica*, vol. 37, no. 2, pp. 285–293, 2009.
- [26] J. B. Rocha-Junior, A. Vlachou, C. Doukeridis, and K. Nørkvåg, "Efficient execution plans for distributed skyline query processing," in *Proceedings of the 14th International Conference on Extending Database Technology: Advances in Database Technology (EDBT '11)*, pp. 271–282, Uppsala, Sweden, March 2011.
- [27] Y. Yongtao and W. Yijie, "Towards estimating expected sizes of probabilistic Skylines," *Science China*, vol. 53, no. 1, pp. 1–18, 2010.
- [28] B. Cui, H. Lu, Q. Xu, L. Chen, Y. Dai, and Y. Zhou, "Parallel distributed processing of constrained skyline queries by filtering," in *Proceedings of the 24th International Conference on Data Engineering (ICDE '08)*, pp. 546–555, IEEE Computer Society, Cancun, Mexico, April 2008.
- [29] X. Ding and H. Jin, "Efficient and progressive algorithms for distributed skyline queries over uncertain data," in *Proceedings of the 30th IEEE International Conference on Distributed Computing Systems (ICDCS '10)*, pp. 149–158, Genoa, Italy, June 2010.

## Research Article

# Transit Station Congestion Index Research Based on Pedestrian Simulation and Gray Clustering Evaluation

**Shu-wei Wang, Li-shan Sun, Jian Rong, and Zi-fan Yang**

*Key Laboratory of Transportation Engineering, Beijing University of Technology, Beijing 100124, China*

Correspondence should be addressed to Li-shan Sun; [lssun@bjut.edu.cn](mailto:lssun@bjut.edu.cn)

Received 2 September 2013; Accepted 13 November 2013

Academic Editor: Geert Wets

Copyright © 2013 Shu-wei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A congestion phenomenon in a transit station could lead to low transfer efficiency as well as a hidden danger. Effective management of congestion phenomenon shall help to reduce the efficiency decline and danger risk. However, due to the difficulty in acquiring microcosmic pedestrian density, existing researches lack quantitative indicators to reflect congestion degree. This paper aims to solve this problem. Firstly, platform, stair, transfer tunnel, auto fare collection (AFC) machine, and security check machine were chosen as key traffic facilities through large amounts of field investigation. Key facilities could be used to reflect the passenger density of a whole station. Secondly, the pedestrian density change law of each key traffic facility was analyzed using pedestrian simulation, and the load degree calculating method of each facility was defined, respectively, afterwards. Taking pedestrian density as basic data and gray clustering evaluation as algorithm, an index called Transit Station Congestion Index (TSCI) was constructed to reflect the congestion degree of transit stations. Finally, an evaluation demonstration was carried out with five typical transit transfer stations in Beijing, and the evaluation results show that TSCI can objectively reflect the congestion degree of transit stations.

## 1. Introduction

Urban rail transit is one of the most important public transport modes. Transit stations, especially transfer stations, suffer from large passenger flow pressure during peak hours, and congestion phenomenon often occurs. It is easy to understand that congestion phenomenon in transit transfer stations could lead to low transfer efficiency, as well as hidden danger of passenger security. Thus, effective management of congestion phenomenon shall help to reduce the efficiency decline and danger risk. However, due to the difficulty in acquiring microcosmic pedestrian density, most of the existing researches lack quantitative indicators to reflect congestion degree. In order to quantify congestion degree, this paper puts forward a concept of Transit Station Congestion Index (TSCI) according to Global Port Congestion Index (GPCI). GPCI is published weekly to detailedly and timely reflect the retention situation of coal, ore, and other dry bulk fleets in major ports around the world, for the purpose of analyzing the influence of port congestion on the supply and demand of dry bulk

market. Up to now, GPCI has covered 80 major ports from different countries, including Australia, Brazil, China, India, and South Africa. Just like GPCI, TSCI could help to achieve effective control of transit station congestion phenomenon, which would help to reduce the efficiency decline and danger risk.

Previous researches mainly focus on the qualitative analysis of traffic facility pedestrian flow characteristics, which can be outlined as below. Older [1] found that the pedestrian flow characteristics of different traffic facilities vary a lot. Helbing [2] found that the geometric boundary of traffic facilities can affect pedestrian space and facility capacity. Fruin [3] put forward a traffic facility level of service evaluation and classification theory based on the pedestrian density, through the analysis of stair, platform, and traffic facilities' pedestrian density characteristic. Based on the research of Fruin, the United States Transportation Research Board [4] proposed a pavement level of service classification standard in the Highway Capacity Manual (HCM). Seneviratne and Morrall [5] and Tanaboriboon et al. [6] found that pedestrian traffic

characteristics of different countries vary greatly and pointed out that all traffic facility researches should be combined with local pedestrian characteristic.

In recent years, some researchers began to study the microbehavior law of public transport pedestrian. Hoogenboom et al. proposed a NOMAD model to describe pedestrian behavior and a SimPed model to describe pedestrian flow characteristics under the influence of the interacting of different public transport means [7–10]. Kitazawa and Batty modified the profit condition of shortest path choices under the influence of uncertain factors and conducted a research of the pedestrian movement path with a genetic algorithm [11]. Schelhorn et al. constructed a STREETS model through the analysis of individual characteristics and attraction points in pedestrian movement and described the movement situation of urban streets using the Agent model [12]. Desyllas and Duxbury analyzed the influence of different street patterns on pedestrian attraction through the method of modeling city morphology and quantified the influence of city infrastructure on pedestrian flow volume with the visual analysis method [13]. Kwon et al. researched in the influence of narrow city streets on the pedestrians detour obstacles behavior [14]. Hine evaluated the influence of traffic environment on behavior and safety perception of pedestrians [15]. Peter Thompson classified and described the pedestrian behavior under building evacuation circumstances and pointed out that pedestrian behavior will be influenced by corridor width and mentation [16]. Gipps and Marksjö [17] developed a cellular automaton model to analyze the pedestrian motion law in a continuous space. Hoofendoorn [18] conducted further research on pedestrian simulation based on the hypothesis of individual utility maximization and proposed a pedestrian route choice and activity planning model.

In general, most of the current researches focus on individual problems and lack quantitative indicators to reflect transit station congestion degree. This paper aims to solve the above problems and an index called Transit Station Congestion Index (TSCI) was constructed to reflect the congestion degree of transit stations. Firstly, in order to quantify congestion degree, the density change law of key traffic facilities in a transit station was analyzed through pedestrian simulation. Secondly, the load degree calculating method of each kind of key traffic facility was defined based on its pedestrian density change law. Thirdly, taking pedestrian density as basic data and gray clustering evaluation as algorithm, an index called Transit Station Congestion Index (TSCI) was constructed to reflect the congestion degree of transit stations. Finally, the practicability of TSCI was verified with an evaluation demonstration of five typical transit transfer stations in Beijing.

This paper consists of six parts: the first part summarized the existing researches related to transit station congestion degree; the second part determined the key traffic facilities in a transit transfer station and conducted pedestrian simulation to analyze the pedestrian density change law of each key traffic facility; the third part proposed the load degree calculating method of each key traffic facility on the basis of its pedestrian density change law; the fourth part constructed an index called Transit Station Congestion Index (TSCI) to reflect

TABLE 1: Transfer form and transfer line of typical stations.

Transit station	Transfer lines	Transfer form
SHD	Line Batong and Line 1	Station hall transfer
FXM	Line 1 and Line 2	Tunnel transfer
GM	Line 1 and Line 10	Tunnel transfer
HXXJNK	Line 5 and Line 10	Station hall transfer
SJZ	Line 5 and Line Yizhuang	Platform transfer

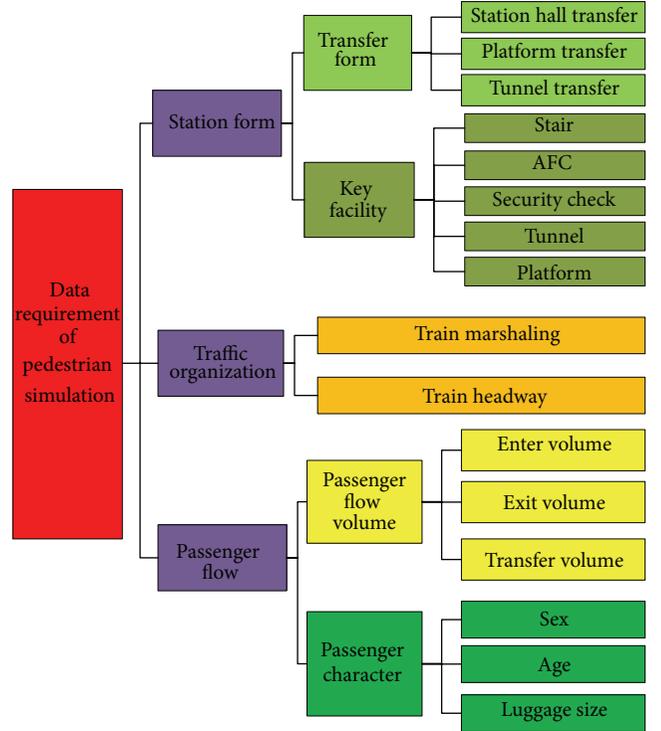


FIGURE 1: Factors affecting transit station congestion degree.

the congestion degree of transit stations by taking pedestrian density as basic data and gray clustering evaluation as algorithm; the fifth part carried out an evaluation demonstration of five typical transit transfer stations in Beijing; the last part is conclusions of this paper.

## 2. Pedestrian Simulation

*2.1. Field Investigation.* The congestion degree of transit stations is typically affected by three factors: station form, traffic organization, and passenger flow; these three types of data are also the data requirements of pedestrian simulation, as shown in Figure 1.

Considering both the diversity of transfer form and the severity of congestion, SHD station, FXM station, GM station, HXXJNK station, and SJZ station were chosen as typical stations to conduct research; their transfer form and transfer lines are shown in Table 1.

In order to conduct pedestrian simulation, field investigation of selected transit stations needed to be carried out.



FIGURE 2: Part of the positions of passenger flow investigation.

TABLE 2: Proportion of different types of passengers.

Transit station	No luggage ratio	Small luggage ratio
SHD	90%	10%
GM	85%	10%
HXXJNK	85%	15%
FXM	80%	20%
SJZ	80%	20%

The investigation time was chosen as the morning peak period (07:00–09:00) of July 2, 4, and 6, 2012. As mentioned in Figure 1, the survey content included station form, traffic organization, and passenger flow, where the passenger flow investigation was conducted by manual counting, as shown in Figure 2.

**2.2. Simulation Model Construction.** There is plenty of pedestrian simulation software to choose from, including LEGION, STEPS, BUILDINGEXDOUS, MYRIAD, PAXPORT, and SIMULEX. Each of them has its speciality and should be selected according to research purpose. Research purpose of this paper is to analyze the pedestrian density change law of different facilities in a transit station of Beijing; therefore, considering LEGION has a pedestrian motion model which was established based on the characteristic of Beijing citizens and at the same time has a well-designed simulation model for each kind of traffic facilities in a transit transfer station, it was eventually chosen. The construction procedure of pedestrian model includes the following: draw the CAD simulation map based on station form, import simulation map into LEGION and add transfer facilities, set traffic organization scheme, and finally add passenger flow information. Choose “Chinese Commuters” as pedestrian type and the passenger character is shown in Table 2.

Part of the other key parameters includes the following.

- (i) Escalator: 0.65–0.75 M/s; tilt angle 30–45°.
- (ii) AFC: delay time as 2 s.
- (iii) Security check: reception capacity as 3–5 persons at a time; delay time 2.25 s.

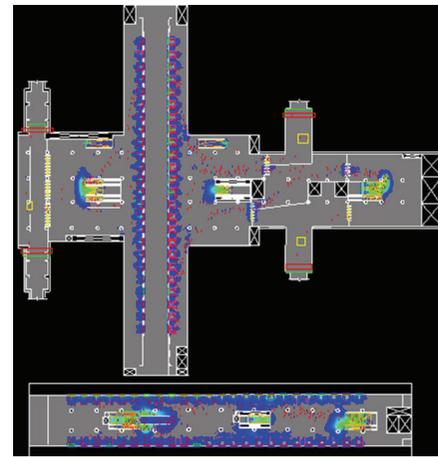


FIGURE 3: Simulation model of HXXJNK station.



FIGURE 4: Simulation model of SHD station.

The complete simulation models are shown as Figures 3, 4, 5, 6, and 7.

**2.3. Key Traffic Facility Pedestrian Density Analysis.** Transit stations are usually too complicate, and it is difficult to analyze the pedestrian density of all areas. However, there are typical traffic facilities in every transit station, whose capacity usually

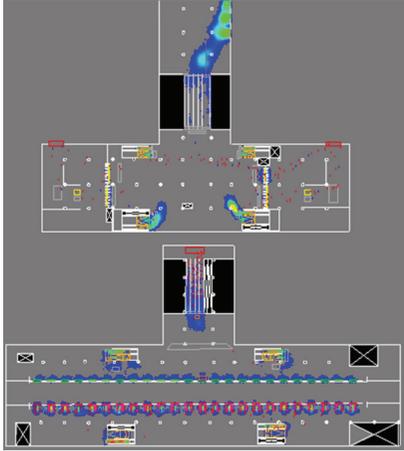


FIGURE 5: Simulation model of SJZ station.

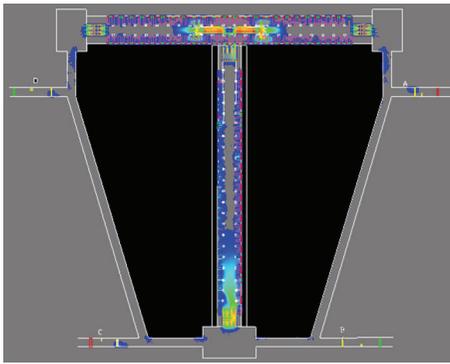


FIGURE 6: Simulation model of FXM station.



FIGURE 7: Simulation model of GM station.

makes them bottlenecks of passenger flow lines, such as security check, AFC, stair, tunnel, ticket office, and platform. Field investigation shows that platform, stair, tunnel, security and AFC are traffic facilities with the highest utilization rate and could reflect the pedestrian density characteristics of the whole station. Therefore, choose these five facilities as key traffic facilities in a transit station and analyze their pedestrian density change law. The pedestrian density data of the objects of study per 0.6 second during simulation period are shown as Figures 8, 9, 10, 11, and 12.

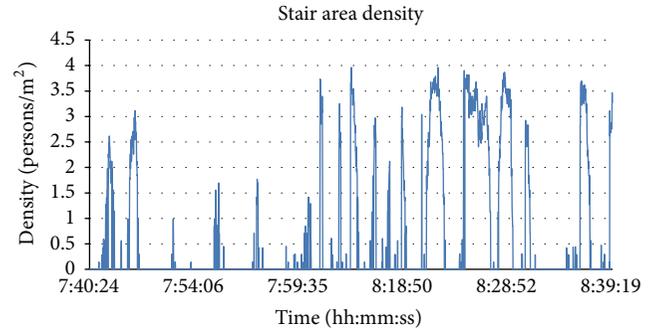


FIGURE 8: Average pedestrian density of stair area every 0.6 second.

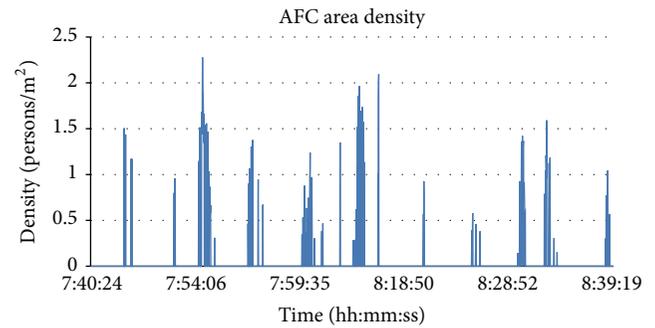


FIGURE 9: Average pedestrian density of AFC area every 0.6 second.

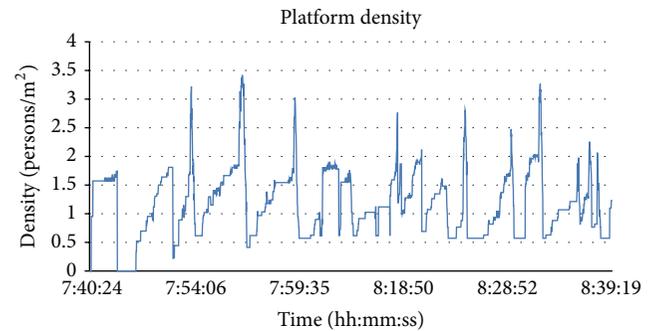


FIGURE 10: Average pedestrian density of platform area every 0.6 second.

### 3. Calculating Method of Key Traffic Facility Load Degree

The pedestrian density of key traffic facilities changes periodically, repeating the process from passenger arriving to dissipation. In order to get the density variation range of each facility, define the load degree when pedestrian density is zero as minimum load and maximum load when pedestrian density reaches the summit. Given the fact that the pedestrian change laws of each type of traffic facility vary from each other, their load degree calculating standards should be different. Based on the above analysis, remove the data where pedestrian density is zero, and rank the rest in an ascending order. Ranking result shows that the passenger density data

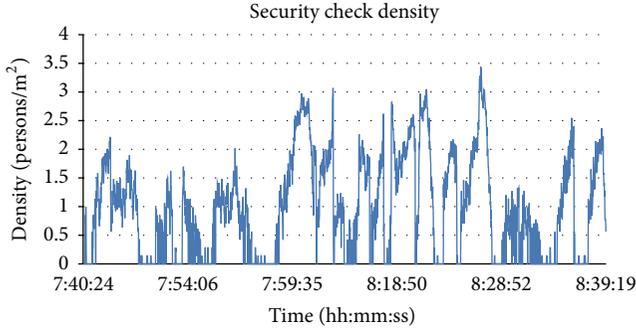


FIGURE 11: Average pedestrian density of security check area every 0.6 second.

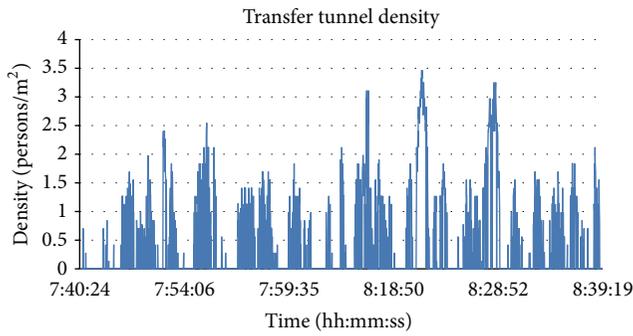


FIGURE 12: Average pedestrian density of transfer tunnel every 0.6 second.

of each key traffic facility in a typical period obeys a normal distribution. As we all know, normal distribution has two special points on the midline symmetry, which are called 15% fractile and 85% fractile; the pedestrian density growth rate between these two points is significantly higher than that of other regions. Take the pedestrian density at the 15% fractile, 50% fractile, and 85% fractile of each type of traffic facility as the load degree calculating standard. Therefore, load degree is divided into four categories, as shown in Table 3, and the unit is passengers per square meter.

#### 4. Grey Clustering Evaluation Model of Transit Station Congestion Index

The passenger flow status in a transit station is rather complex and can be treated as a system in which part of the information is not clear; therefore, chose the pedestrian density of platform, stair, AFC, security check, and tunnel as basic data and grey clustering model as evaluation method to construct a hybrid index, and name this index “transit station congestion index.” The concept of grey system is proposed by Chinese professor Julong Deng and is designed to analyze systems in which part of the information is not clear; in a grey system, all the random variables and stochastic processes are considered as grey variables and grey processes [16]. Grey clustering method is a method based on the establishment of whitenization weight function; it clusters objects according to

TABLE 3: Density value of traffic facility load degree.

Load degree	Facility				
	Security check	AFC	Tunnel	Stair	Platform
I	0.148	0.713	0.612	0.982	0.424
II	0.349	1.466	1.061	2.617	1.056
III	0.819	1.956	1.804	3.607	2.321
IV	2.915	3.487	3.961	5.006	4.867

the whitenization weight function of different indexes, summarizes these indexes into grey categories, and determines which grey category the clustering object belongs to. According to the load degree classification standard in Table 3, establish the whitenization weight function of each type of traffic facility:

$$\begin{aligned}
 f_j^1(x_{ij}) & \begin{cases} 1, & x_{ij} = S_{j1}, \\ \frac{x_{ij} - S_{j2}}{S_{j1} - S_{j2}}, & S_{j1} < x_{ij} < S_{j2}, \\ 0, & x_{ij} \geq S_{j2}, \end{cases} \\
 f_j^2(x_{ij}) & \begin{cases} 1, & x_{ij} = S_{j2}, \\ \frac{x_{ij} - S_{j1}}{S_{j2} - S_{j1}}, & S_{j1} < x_{ij} < S_{j2}, \\ \frac{x_{ij} - S_{j3}}{S_{j2} - S_{j3}}, & S_{j2} < x_{ij} < S_{j3}, \\ 0, & x_{ij} \leq S_{j1} \text{ or } x_{ij} \geq S_{j3}, \end{cases} \\
 f_j^3(x_{ij}) & \begin{cases} 1, & x_{ij} = S_{j3}, \\ \frac{x_{ij} - S_{j2}}{S_{j3} - S_{j2}}, & S_{j2} < x_{ij} < S_{j3}, \\ \frac{x_{ij} - S_{j4}}{S_{j3} - S_{j4}}, & S_{j3} < x_{ij} < S_{j4}, \\ 0, & x_{ij} \leq S_{j2} \text{ or } x_{ij} \geq S_{j4}, \end{cases} \\
 f_j^4(x_{ij}) & \begin{cases} 1, & x_{ij} \geq S_{j4}, \\ \frac{x_{ij} - S_{j3}}{S_{j4} - S_{j3}}, & S_{j3} < x_{ij} \leq S_{j4}, \\ 0, & x_{ij} \leq S_{j3}, \end{cases}
 \end{aligned} \tag{1}$$

where  $f_j^1(x_{ij})$ ,  $f_j^2(x_{ij})$ ,  $f_j^3(x_{ij})$ , and  $f_j^4(x_{ij})$  are the whitenization weight functions of the four grey categories of facility  $j$ ,  $S_{j1}$ ,  $S_{j2}$ ,  $S_{j3}$ , and  $S_{j4}$  are the pedestrian density value of the corresponding I, II, III, and IV grey categories of facility  $j$ , and  $x_{ij}$  is the pedestrian density of facility  $j$  of transit station  $i$ .

The pedestrian density characteristics of different facilities vary from each other; if calculated directly, the evaluation

TABLE 4: Dimensionless grey category value.

Grey category	Facility				
	Security check	AFC	Tunnel	Stair	Platform
I	0.051	0.248	0.213	0.341	0.147
II	0.053	0.224	0.162	0.400	0.161
III	0.078	0.186	0.172	0.343	0.221
IV	0.144	0.172	0.196	0.247	0.241

TABLE 5: Clustering weight of each key traffic facility.

Grey category	Facility				
	Security check	AFC	Tunnel	Stair	Platform
I	0.51	0.11	0.12	0.08	0.18
II	0.49	0.12	0.16	0.07	0.16
III	0.41	0.17	0.19	0.09	0.14
IV	0.27	0.22	0.20	0.16	0.16

function of facilities with lower density will be smaller; therefore, conduct a dimensionless treatment to the grey values via the following formula (the results are in Table 4):

$$y_{jk} = \frac{S_{jk}}{\sum_1^4 S_{jk}}. \quad (2)$$

Facilities with larger grey category value are more likely to encounter high density and should be given a relatively smaller weight. Therefore, apply the inverse weighting method to determine the clustering weight of each key traffic facility via the following Formula (3) and the results are shown in Table 5:

$$\eta_{jk} = \frac{1/y_{jk}}{\sum_{j=1}^m (1/y_{jk})}, \quad (3)$$

where  $\eta_{jk}$  is the clustering weight of grey category  $k$  of facility  $j$ ;  $y_{jk}$  is the dimensionless value of grey category  $k$  of facility  $j$ .

Calculate the clustering evaluation value of transit station  $i$  in grey category  $t$  via the following Formula:

$$\sigma_i^k = \sum_{j=1}^m f_j^k(x_{ij}) \eta_{jk}. \quad (4)$$

Therefore, the grey clustering evaluation sequence of transit station  $i$  is  $\sigma_i = \{\sigma_{i1}, \sigma_{i2}, \sigma_{i3}, \sigma_{i4}\}$ , and the grey category it belongs to should be  $\sigma_i^* = \max\{\sigma_{i1}, \sigma_{i2}, \sigma_{i3}, \sigma_{i4}\}$ . Grey category reflects the congestion degree of a transit station, and the higher the grey category, the higher the congestion degree.

## 5. Evaluation Demonstration

Record the pedestrian density (including zero) of the five typical transit stations during peak hour (07:30–08:30), rank

TABLE 6: Initial index value of evaluation demonstration.

Transit station	Facility				
	Security check	AFC	Tunnel	Stair	Platform
SHD	2.14	1.963	1.34	3.287	4.409
GM	0.594	1.98	1.839	3.614	1.923
HXXJNK	0.31	1.553	2.192	3.395	1.813
FXM	0.304	2.012	1.358	3.932	1.641
SJZ	0.565	1.85	2.193	2.622	1.819

TABLE 7: Initial index value of evaluation demonstration.

Results	Facility				
	SHD	GM	HXXJNK	FXM	SJZ
I	0.00	0.00	0.10	0.11	0.00
II	0.12	0.29	0.57	0.57	0.42
III	0.48	0.75	0.34	0.38	0.56
IV	0.30	0.01	0.04	0.04	0.04
Grey category	III	III	II	II	III

them in ascending order, and take the 85% fractile value as the initial index value, as shown in Table 6, and the unit is passengers per square meter.

Calculate the clustering evaluation value of each station via (4), and the results are shown in Table 7.

As shown in Table 7, SHD, GM, and SJZ stations belong to grey category III, and their congestion degree is relatively high; HXXJNK and GXM stations belong to grey category II, and the congestion degree is relatively low.

Compare the evaluation results with simulation results, and verify the objectivity of the proposed TSCI evaluation method. Simulated average pedestrian density distribution is shown in Figures 13 and 14, where different colors represent different density values, as shown in Figure 15.

Pedestrian density simulation results show that most areas of GM and SJZ stations stay between  $3.59 \text{ P/m}^2$  and  $5.38 \text{ P/m}^2$ , few areas above  $5.38 \text{ P/m}^2$ , and the congestion degree is relatively high; most areas of HXXJNK and FXM stations stay between  $1.08 \text{ P/m}^2$  and  $1.54 \text{ P/m}^2$ , very few areas above  $5.38 \text{ P/m}^2$ , and the congestion degree is relatively low.

According to the above analysis, simulation results and evaluation conclusion are consistent, which means the proposed evaluation method of TSCI is feasible.

## 6. Conclusion

Through the pedestrian density analysis conducted in this paper, the load degree calculating method of each key traffic facility was proposed, which could help to lay a basis for the quantitative analysis of transit station passenger flow. Additionally, the construction of Transit Station Congestion Index (TSCI) is a beneficial attempt to reflect the congestion degree of transit stations, and the evaluation demonstration

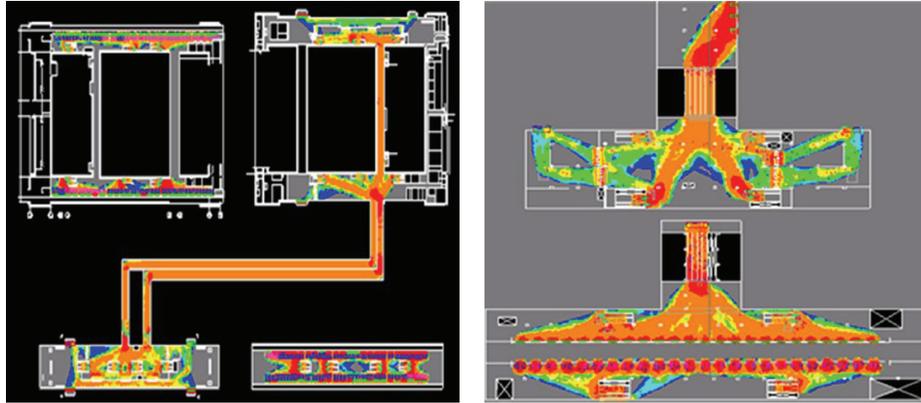


FIGURE 13: Simulated average pedestrian density distribution of GM and SJZ stations (TSCI = 3).

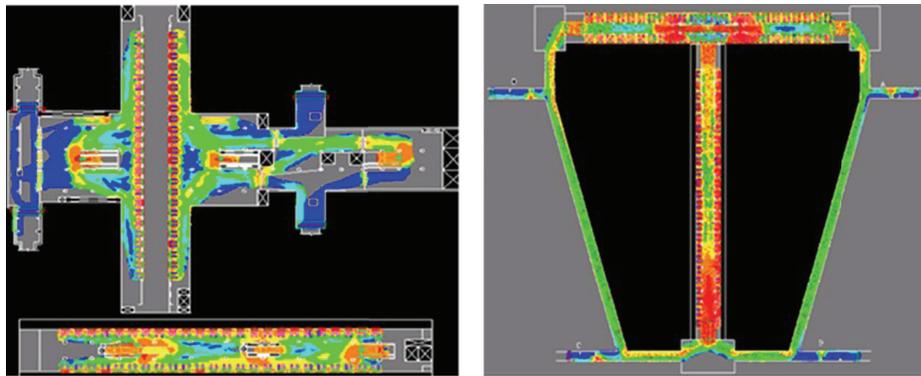


FIGURE 14: Simulated average pedestrian density distribution of HXXJNK and FXM stations (TSCI = 2).

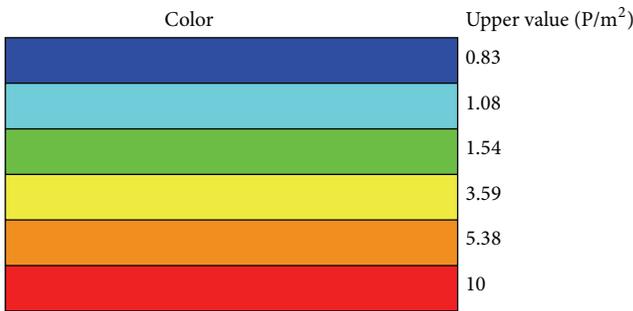


FIGURE 15: Control chart of simulated color and pedestrian density value.

shows that TSCI could reflect the congestion degree of transit transfer stations objectively.

Research findings could act as references for the design and construction of new transit stations, as well as instructions for the improvement of existing ones.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (no. 51308017), the Beijing Municipal Natural Science Foundation (no. 8122009), Specialized Research Fund for the Doctoral Program of Higher Education (no. 20121103120025), the China Scholarship Council, and the Ministry of Industry and Information Technology of China under the Major Program of National Science and Technology with no. 2013ZX01045003-002.

### References

- [1] S. J. Older, *Movement of Pedestrians on Footways in Shopping Streets*, vol. 10, Traffic Engineering & Control, London, UK, 1968.
- [2] D. Helbing, "A fluid-dynamic model for the movement of pedestrians," *Complex Systems*, vol. 6, no. 5, pp. 391–415, 1992.
- [3] J. J. Fruin, *Pedestrian Planning and Design*, Metropolitan Association of Urban Designers and Environmental Planners, New York, NY, USA, 1971.
- [4] *Transportation Research Board: Highway Capacity Manual*, National Research Council, Washington, DC, USA, 1994.
- [5] P. N. Seneviratne and J. F. Morrall, "Analysis of factors affecting the choice of route of pedestrians," *Transportation Planning & Technology*, vol. 10, no. 2, pp. 147–159, 1985.

- [6] Y. Tanaboriboon, S. S. Hwa, and C. H. Chor, "Pedestrian characteristics study in Singapore," *Journal of Transportation Engineering*, vol. 112, no. 3, pp. 229–235, 1986.
- [7] S. P. Hoogendoorn, "Normative pedestrian flow behaviors," LVV rapport, VK 2001.002, Transportation and Traffic Engineering Section, Delft University of Technology, Delft, The Netherlands, 2001.
- [8] S. P. Hoogendoorn, "Pedestrian flow modeling by adaptive control," in *Proceedings of the 83th Transportation Research Board Annual Meeting*, Washington, DC, USA, 2004.
- [9] S. P. Hoogendoorn and P. H. L. Bovy, "Normative pedestrian behavior theory and modeling," in *Proceedings of the 15th International Symposium on Transportation and Traffic Theory*, pp. 219–245, Adelaide, Austral, July 2002.
- [10] S. P. Hoogendoorn, "Pedestrian travel behavior in walking areas by subjective utility optimization," in *Proceedings of the 81th Transportation Research Board Annual Meeting*, Washington, DC, USA, January 2002.
- [11] K. Kitazawa and M. Batty, "Pedestrian behaviour modelling: an application to retail movements using a Genetic Algorithm," in *Proceedings of the 7th International Conference on Design & Decision Support Systems in Architecture and Urban Planning*, Eindhoven, the Netherlands, July 2004.
- [12] T. Schelhorn, D. O'Sullivan, M. Haklay, and M. Thurstain-Goodwin, "STREETS: an agent-based pedestrian model," in *Proceedings of the 6th International Conference Computers in urban planning and urban management on the edge of the millenium (CUPUM '99)*, FrancoAngeli, Venice, Italy, 1999.
- [13] J. Desyllas and E. Duxbury, "Planning for movement: measuring and modelling pedestrian flows in cities," in *Proceedings of the Royal Institution of Chartered Surveyors (RICS '00) Conference*, London, UK, 2000.
- [14] Y.-I. Kwon, S. Morichi, and T. Yai, "Analysis of pedestrian behavior and planning guidelines with mixed traffic for narrow urban streets," in *Proceedings of the 77th Transportation Research Board*, Washington, DC, USA, 1998.
- [15] J. Hine, "Pedestrian travel experiences: assessing the impact of traffic on behaviour and perceptions of safety using an in-depth interview technique," *Journal of Transport Geography*, vol. 4, no. 3, pp. 179–199, 1996.
- [16] A. Peter Thompson, *Developing new techniques for modeling crowd movement [Ph.D. thesis]*, University of Edinburgh, Edinburgh, UK, 1994.
- [17] P. G. Gipps and B. Marksjö, "A micro-simulation model for pedestrian flows," *Mathematics and Computers in Simulation*, vol. 27, no. 2-3, pp. 95–105, 1985.
- [18] S. P. Hoofendoorn, "Wayfinding under uncertainty in continuous time and space by dynamic programming," in *Proceedings of the 13th Mini-EURO Conference*, vol. 6, pp. 71–76, 2002.

## Research Article

# Performance of the Priority Control Strategies for Bus Rapid Transit: Comparative Study from Scenario Microsimulation Using VISSIM

Min Yang,<sup>1</sup> Wei Wang,<sup>1</sup> Bo Wang,<sup>2</sup> and Jing Han<sup>1</sup>

<sup>1</sup> School of Transportation, Southeast University, No. 2, Sipailou, Nanjing 210096, China

<sup>2</sup> Department of Civil and Environmental Engineering, Texas A&M University, 1226 TAMU, College Station, TX 77843-1226, USA

Correspondence should be addressed to Min Yang; transtar2002@163.com

Received 22 September 2013; Accepted 5 November 2013

Academic Editor: Huimin Niu

Copyright © 2013 Min Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bus rapid transit (BRT) has a great potential to improve the service level of transit system and has been implemented in many Chinese cities. However, the priority it can provide to buses has not been explored fully. Therefore, this study mainly investigated two advanced control strategies (signal priority using advanced detection and transit speed control). Signal priority using advanced detection is a strategy which detects one cycle ahead of buses' arrival in order to adapt a more flexible control algorithm to provide signal priority for buses. Another is transit speed control, which provides priority at intersections for buses by controlling the speed of them and predicting their arrival at certain intersection. These two advanced strategies were modeled and evaluated using simulation software VISSIM and presented better performance than other three scenarios (base case, exclusive bus lane, and conventional transit signal priority). Only the eastbound direction would be researched as its traffic flow and bus volume are much larger than those of the other direction. Data used in this model was collected in Yingtan City. It is also shown that both the operation of BRT and the efficiency of private traffic can be much improved by applying the two strategies proposed above.

## 1. Introduction

Bus rapid transit (BRT) is a high-quality bus-based transit system that delivers fast, comfortable, and economical urban mobility through the provision of segregated right-of-way infrastructure, rapid and frequent operations, and excellence in marketing and customer service [1, 2]. Compared with the traditional rapid transit system (such as rapid rail transit), BRT system requires lower initial capital investment and shorter implementation time and provides more flexible routes. Compared to normal bus transit (NBT), it has larger capacity, better passenger facilities such as exclusive bus lanes, fare collection system, real-time information system, and modern bus stations. It also combines advantages such as the flexibility of conventional buses and the operational efficiency of rail transit [3, 4]. As a promising alternative, BRT has existed for almost 40 years, while it has not been introduced into China until the late 20th century. Cai and Xu discussed BRT application forms and the urgency of popularizing BRT

in China [5]. To accommodate to increasing travel demand, many cities such as Beijing, Hangzhou, Changzhou, and Jinan have developed BRT system; also many other cities including the city of Yingtan mentioned in this study are planning to implement BRT system.

However, priorities that BRT systems in China provide for buses are not fully exploited. Mostly, exclusive bus lanes and conventional transit signal priority are implemented, which limits the efficiency of BRT [6, 7]. Moreover, heterogeneous traffic conditions (prevailing in Chinese cities) will limit the efficiency of the implementation of exclusive bus lanes [8]. Thus, deeper research is needed to further investigate control strategies that can be integrated with BRT. To this end, this study tries to improve BRT efficiency using advanced traffic control strategies.

Transit signal priority (TSP) has been pervasively implemented worldwide, which includes passive, active, and adaptive priority treatment [9]. Passive TSP strategy does not require any detector or request activation. It provides transit

vehicles with signal priority by timing signal plan favoring these vehicles. In general, when transit operations are predictable with stable routes and arrival schedules, passive priority strategies can be an efficient form of TSP [10].

Active TSP needs the implementation of detection and responds to priority requests generated by transit vehicles. Green extension strategy and early green strategy are two well developed and widely used active TSP strategies. Active TSP also can be implemented as unconditional and conditional. Unconditional TSP provides priority for each transit vehicle that sends a request. Conditional TSP offers priority for vehicles only under certain conditions. Previous studies have shown that the net impact of active TSP in terms of delay can be positive or negative, depending on factors such as signal timings, travel demand, and TSP strategy parameters [11]. Also the location of bus-only lane can impact the performance of TSP strategy. Simulation study showed that curb bus-only lane arrangement appeared to have an advantage over median bus-only lanes arrangement in improving overall intersection performance if green extension and early green were deployed [12].

Adaptive TSP considers the tradeoffs between transit and traffic delay and optimizes signal timing plan dynamically through real-time detection. However, currently there is a lack of comprehensive documentation of the effectiveness of such transit priority measures over a wide range of traffic levels, network configuration, technology sophistication and bus volume, and transit frequency and characteristics [13].

Many researchers agreed that TSP can work better under certain conditions such as phase length limitation [14, 15]. But, in this case, conditional TSP strategies may ignore some transit priority calls due to this limitation and thus fail to provide signal priority to every bus and increase bus delay. To address this problem, Wadjas and Furth proposed a new TSP strategy using advanced detection to provide signal priority for light rail trains [16]. Instead of relying on detection only a few seconds in advance of the stop line, they developed a control algorithm in which trains are detected two to three cycles in advance of their arrival at intersection. Then through active modification of signal timing plan in advance, all buses can receive preference treatment.

Other than signal control, speed control strategies also can offer priority to buses. Wang et al. proposed a transit speed control strategy to dynamically control the operating speed of transit to make sure that it can arrive at intersections within certain time range so that preference treatment can be obtained [17]. They also recommended a near-side bus stop design at each intersection to accommodate to dwelling buses. This type of design requires buses to drive at a particular speed and arrive at bus stops right in front of intersections during red interval. After serving on-offs during the red interval and when the signal turns green, buses can drive through intersections without being delayed (dwell time does not count as control delay).

To offer priority to bus, there are also many other alternative operation plans like dynamic scheduling and deadheading strategies. However, considering the feasibility of these strategies in China and the maneuverability in simulation, these strategies were not taken into account.

Therefore, to evaluate these two advanced control strategies and to find out whether they really work, we modeled these two strategies using VISSIM in our study together with other three scenarios (base case, bus lane, and conventional TSP) as comparison. We hope to figure out the genuinely effective strategies, so that the efficiency of BRT could be improved by a large margin.

This paper is organized as follows. In Section 2, the Yingtan city and the model preparation of Shengli avenue are briefly introduced. In Section 3, two proposed control strategies and other three conventional ones are explained. In Section 4, simulation results of those five scenarios are evaluated and analyzed. Finally, summary and recommendations are made.

## 2. Simulation Preparation

In order to analyze potential of the proposed strategies, a behavior-based microscopic simulation software VISSIM was used in this study. With simulation software, we can model these scenarios and analyze their advantages and disadvantages using a set of indicators.

Indicators used to evaluate proposed strategies' influences on transport system generally include three sets: (1) impacts on bus efficiency and reliability (bus delays, travel speed, and travel time consistency), (2) impacts on private traffic (automobile delays and travel speed), and (3) overall impacts on whole network (person delay) [18]. Indicators adopted here include the following:

- (i) bus delay and travel speed,
- (ii) bus reliability,
- (iii) automobile delay and travel speed,
- (iv) average person delay.

Using VISSIM simulation software, we modeled Shengli avenue, which is an important corridor in the downtown area of Yingtan. Located in the northeast of Jiangxi province, Yingtan serves as the transportation junction from central to southern China. In 2009, the population of downtown area in Yingtan has reached the number of 208,000, GDP per person has exceeded 20,000 RMB, and the number of automobiles was 34,310. The area modeled in this study is a traffic arterial crossing through the downtown area of Yingtan from east to west, along which many commercial and entertaining buildings were constructed, such as Xinhua bookstore, central square, and no.1 primary school of Yingtan (as is shown in Figure 1).

Shengli avenue is a four-lane dual carriageway with three signalized intersections including Zhanjiang, Jiaotong, and Zhengda, and each lane has a width of 3.5 meters. Eastbound bus volume during peak hour is 60 vehicles per hour, and passenger volume of maximum section is 3000 persons per hour. Yingtan is at the middle of a rapid developing period, during which travel demand is increasing dramatically. To relieve traffic congestion, Yingtan has planned to provide priority to public transit and recently is about to install a BRT lane along Shengli avenue (shown in Figure 2).

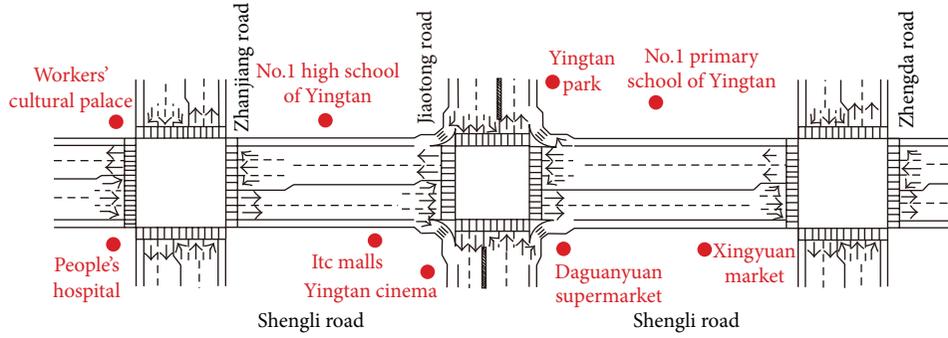


FIGURE 1: Current intersections and land use along Shengli road.

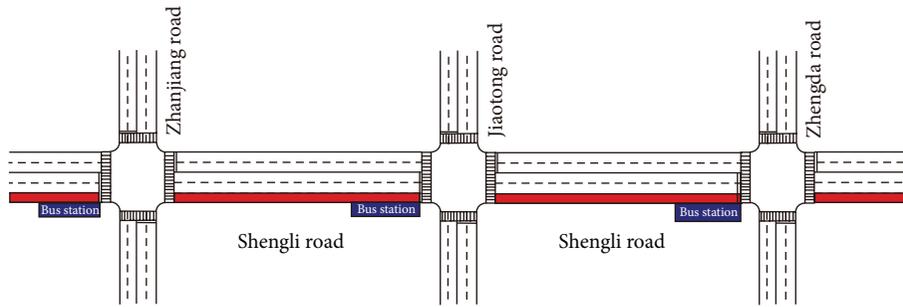


FIGURE 2: The exclusive lane and bus stops of BRT planned on Eastbound Shengli road.

Besides the implementation of BRT lane, Yingtan also has prepared to install a bus stop in front of stop line at each intersection along this arterial. There are mainly two reasons: one is to facilitate passengers to transfer from one route to another; the other is to make use of the period of red signal to load and unload riders. Thus, the objective of these strategies mentioned in this study may have a slight difference from normal situation as normally priority strategies' objective is to provide green signal for buses at intersections to allow them to pass without stop or delay. On the other hand, in this study, because the bus stops are installed in front of stop lines at intersections, priority strategies are trying to provide red signal for buses to allow them to have sufficient time to load and unload passengers and then drive through intersections during green signal. The capacity of bus stop is calculated using a microscopic model presented by Fernández (2010) [19].

With the help of data provided by the final report on urban transit planning of Yingtan (2011–2020) [20], we modeled Shengli avenue in VISSIM. As the microscopic simulation software VISSIM is driving behavior based [21], parameters concerning driving model are essential for model validation. To establish a behavior model of Chinese driver, this study adopts the parameters adjustment done by Berkhout in his work of parameter calibrating using Chinese driving behavior model [22]. Based on this, we calibrate and validate the model by comparing the observed automobile travel speed to the simulated automobile travel speed from the VISSIM model.

### 3. Simulation Scenarios

This study employed five simulation scenarios including (1) base case, (2) adding exclusive bus lane, (3) conventional active transit signal priority, (4) transit signal priority using advanced detection, and (5) transit speed control. By modeling these five scenarios in VISSIM, we tried to analyze the impacts of two proposed priority strategies (modeled as scenarios (4) and (5)) on public transit as well as on private traffic. For the reason that the traffic flow and bus volume from eastbound direction are much larger than those from the other one, all of the following scenarios only provide priority for eastbound direction for the convenience of comparison.

*3.1. Scenario (1): Base Case.* The first scenario modeled in VISSIM is the current situation of traffic network along Shengli Avenue. This scenario serves as the base case and is used to analyze the impacts that these following treatments would bring to this corridor. In this scenario buses are not given any priority in any form.

*3.2. Scenario (2): Adding Exclusive Bus Lane.* On the basis of scenario (1), scenario (2) adds an eastbound exclusive bus lane on the outer side of the corridor. For our study, only focused on one direction, westbound exclusive bus lane is not installed and priority strategies are not provided for buses driving this direction.

**3.3. Scenario (3): Conventional Active Signal Priority.** Active priority strategies provide priority treatment to a specific transit vehicle following detection and subsequent priority request activation. Traditionally, this kind of strategies relies on detection only a few seconds' distance of travel in front of the stop line. There are two kinds of active strategies implemented in Scenario (3), which are green extension strategy and early green strategy. A green extension strategy extends the green time for the TSP movement when a TSP-equipped vehicle is approaching. This strategy only applies when the signal is green for the approaching buses. An early green strategy shortens the green time of preceding phases to expedite the return to green (i.e., red truncation) for the movement where a TSP-equipped vehicle has been detected. This strategy only applies when the signal is red for the approaching bus vehicle [9].

In simulation model of this scenario, we provide bus with priority by adding a separate lane and adopting TSP strategy described above. Also, two detectors are used for this scenario, one is installed 30 meters in front of the stop line and the other is installed right after the stop line. The first detector generates priority request and second detector is used to determine when the green extension should end.

**3.4. Scenario (4): Active Signal Priority Using Advanced Detection.** This priority strategy was proposed by Wadjas and Furth [16]. Through advanced detection and prediction, this strategy algorithm could begin to adjust signal timing ahead of the arrival of bus; thus there would be more room for signal timing adjusting than simple active signal priority (i.e., scenario (3)). Better flexibility for adjusting leads to less amount of adjustment for each phase, and because of this, the impacts on private traffic would be negligible. Brief process of this strategy is shown in Figure 3.

There are mainly two differences in scenario (4) compared with the original one proposed by Wadjas and Furth [16].

(i) Prediction window is cancelled. Because of the advanced detection used in original strategy, prediction was needed to estimate how much time certain vehicle would need to drive from the detection to stop line. In original strategy, Wadjas proposed a twenty-second long prediction window as many factors would fluctuate transit travel time. However, in scenario (4), these factors do not exist or are too slight to impact bus travel time; implementation of exclusive bus lane protects buses from private traffic. There is no other intersection between detector and its corresponding intersection and bus stops are constructed right in front of the stop line, so dwell time would not fluctuate travel time. Moreover, simulation results show that the standard deviation in this scenario is 2.6 seconds, which proves that bus travel time is very reliable. For these reasons, priority strategy used in scenario (4) cancels prediction window.

(ii) Bus serving signal at intersections is different. Original algorithm allows (if it succeeds) buses to obtain green signal at intersections, while in this scenario,

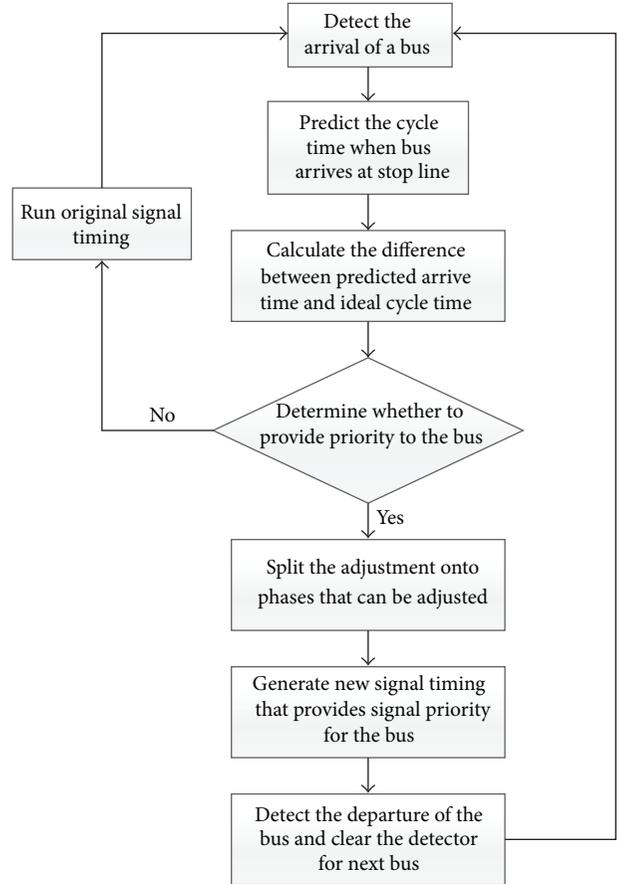


FIGURE 3: Brief algorithm of active signal priority using advanced detection.

as we have discussed, bus serving signal is red at intersections. Together with the bus stops installed in front of the stop lines, this design allows buses to load or unload passengers during period of red signal and then drive through this intersection when it gets green. Target (ideal) arrival time is 20-second red remaining time, which equals the minimum predicted dwell time.

Apart from those two points mentioned above, strategy used in scenario (4) remains the same as original. Figure 4 is an illustration of this algorithm.

As shown in Figure 4, the difference between the predicted arrival time and the desired arrival time is the amount of adjustment. This difference could be eliminated through adjusting signal timing. Because of the advanced detection, the amount of adjustment could be split to those phases among the advance; for example, in Figure 4(a) (compression situation), the adjustable phases could be phases 2, 3, 4, 1, and 2 (phases between the detect time and the desired arrival time). As we can see, adjustment of each phase is much less because the split and impact to the private traffic would be negligible. Similar thing happens in Figure 4(b) (extension situation). In this scenario, all the three signalized intersections along the Shengli Avenue would implement this

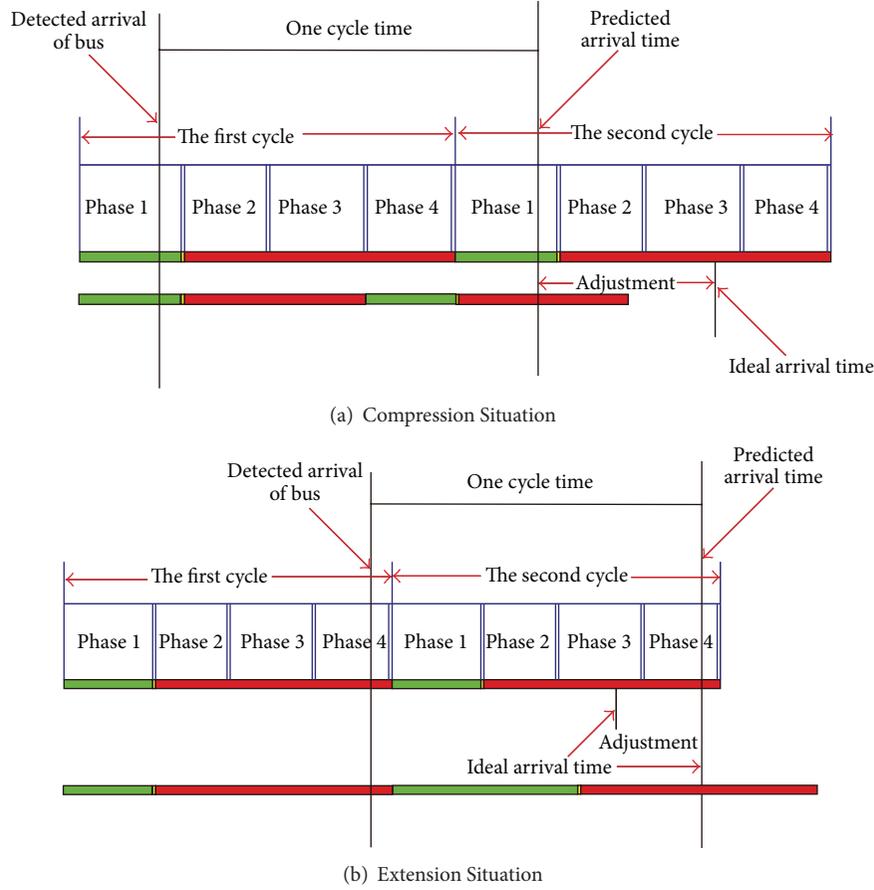


FIGURE 4: Illustrations of Adjustment between the Predicted and Target Arrival Time.

strategy, and each corresponding detector is installed at one cycle's distance of travel ahead of the stop line.

3.5. *Scenario (5): Transit Speed Control.* This strategy is initially proposed by Wang et al. [17], coauthor of this paper. Through guiding and controlling the travel speed, this strategy provides priority for buses at intersections and makes their arrivals at certain intersection predictable. Moreover, there was a theoretical analysis proposed by Wang proving that transit speed control can decrease bus delay. Their analysis is briefly shown below.

Assume that there is an arterial that has  $n$  intersections and  $m$  out of them is installed with near-side bus stops. Considering the base case, in which no bus speed control is adopted and bus stops are installed in the middle of road segment, the travel time of bus running through this arterial can be calculated as follows:

$$\text{total}_1 = \sum_{i=1}^{n+1} \frac{s_i}{v_i} + \sum_{i=1}^n p_i R_i + \sum_{i=1}^m T_i, \quad (1)$$

where

$s_i$  is the length of road segment  $i$ ,  
 $v_i$  is the speed in segment  $i$ ,

$p_i$  is the possibility of a bus stops at intersection  $i$ ,  
 $R_i$  is the stop time of a bus stops at intersection  $i$ , and  
 $T_i$  is the dwell time.

We can see that total travel time in this case consists of three parts: travel time in the road segment, delay at intersection, and dwell time at bus stops. Statistically,  $p_i = 0.5$  and  $R_i = C_{i2}/2$ , where  $C_{i2}$  is the red signal time at intersection  $i$ .

Considering another case, where transit speed control is adopted and bus stops are installed near intersections, the total travel time can be calculated using the following formula:

$$\text{total}_2 = \sum_{i=1}^{n+1} \frac{s_i}{v_i} + \sum_{i=1}^m \max(T_i, C_{i2}), \quad (2)$$

where  $C_{i2}$  is the red signal time at intersection  $i$ .

Assume that the time decrease because of acceleration is approximately equal to the time increase because of deceleration, and  $T_i \leq C_{i2}$ . We have

$$\text{total}_2 \approx \sum_{i=1}^{n+1} \frac{s_i}{v_i} + \sum_{i=1}^m C_{i2}. \quad (3)$$

So by comparing formulations (1) and (3), we can know the time saved by this strategy if there are  $k$  buses running through this arterial as follows:

$$\Delta = k * (\text{total}_1 - \text{total}_2) = k * \left( \frac{1}{4} \sum_{i=1}^n C_{i2} + \sum_{i=1}^m T_i - \sum_{i=1}^m C_{i2} \right). \quad (4)$$

They also did a numerical estimation according to formula (4) to demonstrate how much time this strategy can save. If there are 100 buses, the arterial consists of 10 intersections and 7 of them are installed with near-side bus stops, and dwell time and signal cycle time are 50 s and 60 s, respectively, then  $\Delta = 2.22$  h, which means that the total time saved by this strategy is more than two hours.

However, in this scenario, the strategy is no longer real-time because VISSIM is unable to simulate the real-time speed control.

In order to simulate speed control strategy, we adopt a simplified strategy. Simplification is as follows.

- (i) Equalize signal cycle of three intersections along the corridor.
- (ii) Make the dispatch interval of buses multiples of the equalized cycle.
- (iii) Calculate the desired speed at each portion of road to allow buses to obtain desired signal.

Desired speed is calculated as follows.

Assume that a certain bus departs stop at  $T_1$  time with a speed of  $V$ . This bus needs to travel a distance of  $S$  to arrive at next intersection or stop; then it will take the bus  $T = S/V$  to drive through this portion of road. When it arrives at next intersection, this cycle has run a time of  $T_2 = (T_1 + T)\%C$  (symbol % is used here for calculating residue), so the desired speed in this portion of road would be as follows:

$$V_{\text{opt}} = \frac{S}{T - [T_d - (C - T_2)]}, \quad (5)$$

where  $S$  is the distance to travel to the next intersection or stop,  $C$  is the equalized cycle time, and  $T_d$  is bus dwell time.

If  $T_d - (C - T_2) > 0$ , bus needs to accelerate; otherwise, bus needs to decelerate.

Also, bus cannot accelerate or decelerate too much because of safety and efficiency issues. So Wang has already proposed a speed interval: [15, 40] km/h. If calculated speed exceeds this interval, speed control is cancelled and this bus runs at its original speed.

Here is a demonstration of calculation with numbers to show how the previous equation can be used in real implementation case.

Assume that signal starts at time 00:00:00 and its cycle time is  $C = 100$  s. At time 00:30:00 ( $T_1$ ), a bus dispatched from terminal and detected by detector associated with the first intersection that it will encounter. This bus travels at speed of  $V = 30$  km/h and there is a distance of  $S = 1$  km between his current position and the intersection. Then it will take this bus  $T = 1/30 * 3600 = 120$  s to travel to

the intersection. And when it arrives at the intersection, the current signal cycle has run a time of  $T_2 = (T_1 + T)\%C = (30*60+120)\%100 = 20$  s. Dwell time for this bus is predicted as 30 s. So, recommended speed for this bus would be as follows:

$$V_{\text{opt}} = \frac{S}{T - [T_d - (C - T_2)]} = \frac{1}{120 - [30 - (100 - 20)]} * 3600 = 21.2 \text{ km/h}. \quad (6)$$

All these seemingly meaningless numbers (like 60, 3600) in equations above are used for unit converting.

Because 21.2 km/h lies in the speed interval [15, 40] km/h, which was proposed by Wang et al. (2003), it would be recommended as the desire speed in this segment of road.

In simulation model, buses are running on a separate lane and guided by the speed control algorithm described above.

## 4. Simulation Results

The VISSIM evaluation function makes it possible to the level-of-services for different scenarios. In this section, based on the results of VISSIM models, we evaluate how each priority strategy will impact public transit efficiency and operation of private traffic. Evaluation is going to be analyzed in two categories: buses and automobiles. In the end, we also proposed an indicator of average person delay.

**4.1. Throughput.** Throughput is the ratio of number of trips generated to number of trips completed in each case. Throughput of five scenarios is shown in Table 1.

When the situation is saturated, some parameters should be considered again, such as the speed of vehicles.

**4.2. Buses.** Indicators used in this section to evaluate bus efficiency are delays of buses at intersections, bus travel speed, and bus reliability.

**4.2.1. Delay and Travel Speed.** Delay can be calculated as the time difference between the simulated travel time and the ideal travel time (where vehicle would neither be interrupted by other traffic nor by signals). Delays of buses at three intersections and their average in five scenarios are listed in Table 2, and average travel speed is shown in Table 3.

We can see that compared to scenario (1), delays of buses at intersections in scenario (2) have been decreased dramatically from 46.3 s averagely to 36.2 s and bus travel speed in scenario (2) has been improved from 13.9 km/h to 18.4 km/h, which means that implementation of exclusive bus lanes can improve bus efficiency. However, when adding priority strategies (especially in scenarios (4) and (5)), bus efficiency improvements can achieve a much higher level (average bus delay decreased from 46.3 s to 11.3 s and 9.6 s and bus travel speed increased from 13.9 km/h to 23.8 km/h and 21.0 km/h). These results demonstrate that exclusive bus lanes integrated with priority strategies can improve bus efficiency much more than merely implementing exclusive bus lanes. The reason why the improvement

TABLE 1: Throughput of five scenarios.

Scenario	1	2	3	4	5
Throughput	0.953	0.978	0.978	0.974	0.969

TABLE 2: Bus delays (s) in five scenarios along Eastbound Shengli road.

Scenario section	1	2	3	4	5
Zhanjiang road-Shengli road	36.0	29.9	27.4	11.3	10.3
Jiaotong road-Shengli road	47.0	41.8	30.8	11.4	9.5
Zhengda road-Shengli road	56.0	36.8	27.2	11.2	9.1
Average value	46.3	36.2	28.5	11.3	9.6

TABLE 3: Average bus travel speeds (km/h) in five scenarios along Eastbound Shengli road.

Scenario	1	2	3	4	5
Average speed	21	23.8	19.5	18.4	13.9

of bus efficiency in scenario (3) is less evident than in scenarios (4) and (5) is that active signal priority strategy used in scenario (3) is limited and cannot provide full priority to every detected bus as strategies used in scenarios (4) and (5) do.

4.2.2. *Reliability.* Another important indicator to evaluate bus operation is bus reliability. Bus reliability represents the ability for bus vehicles to arrive at stops according to timetable. Improvement of this indicator can make public transit appeals more to passengers.

To measure bus reliability, we have to know whether the bus arrived at and left the stops solidly. In this paper, the standard deviation of the time that a bus needed to travel from the beginning of Shengli avenue to the end was defined as the quantitative measurement of “bus reliability.”

Simulation results of this indicator are listed in Figure 5 and Table 4.

As shown in Figure 5, this study uses the time that a bus needed to travel from the beginning of Shengli avenue to the end to measure the reliability of buses. The use of travel time to reflect the reliability due to this evaluation data can be directly and easily measured in VISSIM. The fluctuation of these curves in figure above shows the reliability of buses.

We can see that only the curves of scenario (4) and (5) show an excellence of bus reliability, while other three scenarios demonstrate rather fluctuated curves meaning that bus travel time in these scenarios has no satisfactory consistency. These data demonstrate that the implementation of exclusive bus lane and traditional active signal priority contribute little to the improvement of bus reliability. On the other hand, buses have been much more reliable because of strategies like transit signal priority using advanced detection and transit speed control.

4.3. *Automobiles.* Indicators used in this section to evaluate automobile operation are delays of automobiles at intersections and automobile travel speed. Delays and

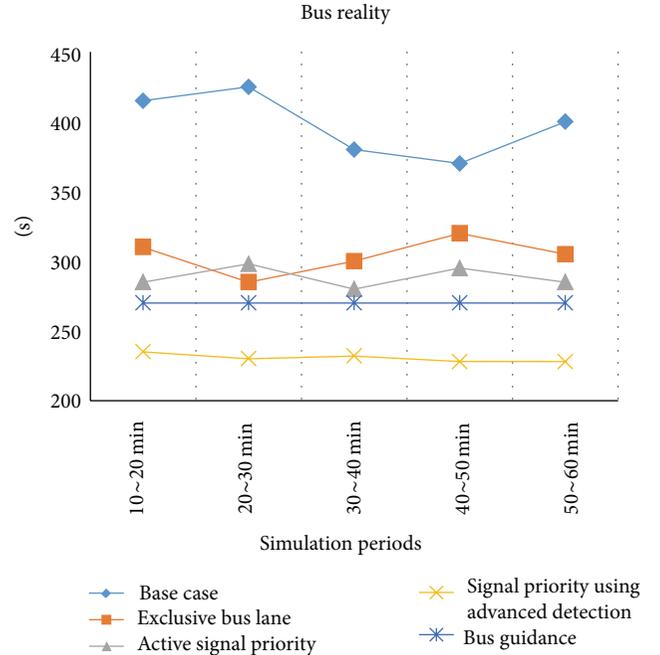


FIGURE 5: Reliability of bus travel time in five scenarios along Eastbound Shengli road.

TABLE 4: Average and standard deviation of bus travel time.

Scenario	1	2	3	4	5
Average (s)	399.5	304.3	290.0	233.8	264.8
Stdev (s)	19.7	11.4	5.9	2.6	1.9

TABLE 5: Car delays (s) in five scenarios along Eastbound Shengli road.

Scenario Section	1	2	3	4	5
Zhanjiang road to Shengli road	15.1	12.5	13.3	13.4	19.2
Jiaotong road to Shengli road	18.2	13.3	15.5	14.2	25.0
Zhengda road to Shengli road	17.4	12.6	16.1	15.1	18.9
Average value	16.9	12.8	15.0	14.2	21.0

TABLE 6: Average car travel speeds (km/h) in five scenarios along Eastbound Shengli road.

Scenario	1	2	3	4	5
Average speed	28.7	33.6	32.2	32.9	26.2

travel speeds of automobiles in five scenarios are shown in Tables 5 and 6.

We can see that adding an exclusive bus lane (scenario (2)) leads to the least delay (averagely 12.8 s) and the fastest travel speed (33.6 km/h), which means this strategy improves private traffic operation conditions at the highest level. On the basis of scenario (2), adding any priority strategies for buses will more or less impact the operation of private traffic in a bad way. What we need to know is which priority strategy will cast the least impacts on private traffic on earth. As we can see in the figures above, private traffic is influenced mostly

TABLE 7: Average person delay.

Scenario	1	2	3	4	5
Person delay (s)	26.7	20.6	16.8	13.3	17.2

in scenario (5) (transit speed control) in which automobile delays have increased averagely from 16.9 s to 21.0 s and travel speed has decreased from 28.7 km/h to 26.2 km/h compared to scenario (1). This is because when implementing nonreal-time speed control strategy in this study (because of the limitation of simulation software) we have to equalize signal cycle time of each intersection which certainly will have major impacts on private traffic. In scenario (4), private traffic operates most fluently (automobile delays have decreased averagely from 16.9 s to 14.2 s and travel speed has increased from 28.7 km/h to 32.9 km/h) because of the advanced detection strategy. As we have discussed before, advanced detection will provide more room for signal timing adjusting which leads to slight adjustment of each phase and impact least on private traffic.

**4.4. Average Person Delay.** In order to evaluate these strategies' impacts, both on public transit and on private traffic, we employ the indicator of average person delay. Average person delay is an indicator that is measured by person instead of vehicle, and it embodies that the priority for buses is actually the priority for persons. Therefore average person delay is an indicator often used in person-oriented projects. Average person delays of these scenarios are shown in Table 7, and we can find that scenarios (4) and (5) improve the system's efficiency mostly.

## 5. Conclusion and Future Directions

In order to seek new ways to improve BRT efficiency, we propose two priority strategies and integrate them with exclusive bus lane to see how they would impact public transport efficiency. These two proposed control strategies are implemented on the basis of BRT features including exclusive bus lanes, bus stops installed in front of stop lines, and bus upload and download passengers during red signal period at intersections. The method we used in this study is microscopic simulation. We modeled these two priority strategies (signal priority using advanced detection and transit speed control) and other three scenarios (base case, exclusive bus lane, and active signal priority) in VISSIM and compared them to analyze how these two priority strategies will impact public transit and private traffic.

Evaluation results show that, in the view of improving bus efficiency, four priority strategies (scenarios (2), (3), (4), and (5)) all have a positive influence on buses, and among those strategies, signal priority using advanced detection and transit speed control have the most positive impacts. Taking the improvement of bus reliability into account, we find that influences of exclusive bus lane and active signal priority are slight, while strategies like priority using advanced detection and speed control remarkably improve the reliability of buses. Considering the impacts on private traffic, we can see that

adding an exclusive bus lane has a positive impact on private traffic; on the other hand, the other three strategies (scenarios (3), (4), and (5)) have negative impacts. However, among those three scenarios, private traffic in scenario (4) bears the least impact, which means that signal priority using advanced detection strategy causes negligible impact on automobiles. From the aspect of average person delay, signal priority using advanced detection improves the system's efficiency mostly, and speed control takes the second place.

When applying the proposed strategies in real world, more complex details should be taken into account, such as the size of intersections, the equipment used to collect data, and how to give the guidance information back to the bus drivers. If the intersection is too big, the time that a bus needs to travel across it and the possible disturbing factors a bus may face will all be worthy of more attention.

This study recommended two priority strategies that can improve BRT efficiency but still ameliorations can be made. In the future, this research can be carried on mainly from these three aspects. First, not only considering traffic flow and bus volume from both the major direction and minor direction in signal priority using advanced detection, but also making out the conflict between prior requests from different directions. Second, developing simulation method that can simulate real-time speed control strategy to test the true potential of this strategy on improving bus efficiency. Third, comparing the conditions where different distances are set for the bus stops from the sections with priority.

In the end, we hope this study will help to push research on public efficiency forward. This work provides a case study for simulation of BRT system in China and probes the ways of improving bus efficiency from the aspect of advanced traffic control strategy.

## Conflict of Interests

The authors declare that they have no conflict of interests to this work. The authors declare that they do not have any commercial or associative interest that represents a conflict of interests in connection with the work submitted.

## Acknowledgments

This research is supported by the National Basic Research 973 Program (2012CB725400) and the National Natural Science Foundation of China (51338003, 51378120, and 51008061). Research funds from Jiangsu Transportation Department (2011Y04) and Foundation for Young Key Teachers of South-east University are also appreciated.

## References

- [1] A. Golub, *Brazil's Buses: Simply Successful*, University of California Transportation Center, Berkeley, Calif, USA, 2004, <http://www.uctc.net/>.
- [2] H. S. Levinson, S. Zimmerman, J. Clinger, and J. Gast, "Bus rapid transit: synthesis of case studies," *Transportation Research Record*, vol. 1841, pp. 1–11, 2003.

- [3] H. S. Levinson, S. Zimmerman, and J. Clinger, "Bus rapid transit-volume 1: case studies in bus rapid transit," Transit Cooperative Research Program (TCRP) Report 90, Federal Transit Administration in Cooperation with the Transit Development Corporation, 2011, [http://www.nbrti.org/docs/pdf/tcrp\\_rpt\\_90v1.pdf](http://www.nbrti.org/docs/pdf/tcrp_rpt_90v1.pdf).
- [4] L. Schramm, K. Watkins, and S. Rutherford, "Features that affect variability of travel time on bus rapid transit systems," *Transportation Research Record*, no. 2143, pp. 77–84, 2010.
- [5] J. C. Cai and K. M. Xu, *Develop the Bus Rapid Transit, Make Changzhou a Better City*, Urban Vehicles, 2008.
- [6] R. B. Diaz, *Characteristics of Bus Rapid Transit for Decision-Making*, U.S. Department of Transportation, McLean, Va, USA, 2004.
- [7] L. Wright and W. Hook, *Bus Rapid Transit Planning Guide*, Institute for Transportation and Development Policy, New York, NY, USA, 3rd edition, 2007.
- [8] V. T. Arasan and P. Vedagiri, "Microsimulation study of the effect of exclusive bus lanes on heterogeneous traffic flow," *Journal of Urban Planning and Development*, vol. 136, no. 1, pp. 50–58, 2010.
- [9] H. R. Smith, *Transit Signal Priority Handbook*, 2005.
- [10] W. J. Ma and X. G. Yang, "A passive transit signal priority approach for bus rapid transit system," in *Proceedings of the 10th International IEEE Conference on Intelligent Transportation Systems (ITSC '07)*, pp. 413–418, Seattle, Wash, USA, October 2007.
- [11] Z. R. Abdy and B. R. Hellinga, "Analytical method for estimating the impact of transit signal priority on vehicle delay," *Journal of Transportation Engineering*, vol. 137, no. 8, pp. 589–600, 2011.
- [12] H. Xu and M. Zheng, "Impact of bus-only lane location on the development and performance of the logic rule-based bus rapid transit signal priority," *Journal of Transportation Engineering*, vol. 138, no. 3, pp. 293–314, 2012.
- [13] A. Skabardonis, "Control strategies for transit priority," *Transportation Research Record*, vol. 1727, pp. 20–26, 2000.
- [14] S. Khasnabis, R. R. Karnati, and R. K. Rudraraju, "NETSIM-based approach to evaluation of bus preemption strategies," *Transportation Research Record*, vol. 1554, pp. 80–89, 1996.
- [15] P. G. Furth and T. H. J. Muller, "Conditional bus priority at signalized intersections: better service with less traffic disruption," *Transportation Research Record*, vol. 1731, pp. 23–30, 2000.
- [16] Y. Wadjas and P. G. Furth, "Transit signal priority along arterials using advanced detection," *Transportation Research Record*, vol. 1856, pp. 220–230, 2003.
- [17] W. Wang, S. Chen, and X. Hu, "Novel integrated method of bus speed guidance and dispatching based on "one route one line and run straight mode"" *Journal of Southeast University*, vol. 38, no. 6, pp. 1110–1115, 2008.
- [18] S. Jiang and M. Murga, "Linking planning and operations for BRT: a microscopic traffic simulation study for the Chicago loop," in *Transportation Research Board (TRB) Meeting*, Washington, DC, USA, January 2010.
- [19] R. Fernández, "Modelling public transport stops by microscopic simulation," *Transportation Research C*, vol. 18, no. 6, pp. 856–868, 2010.
- [20] M. Yang, "Final report on Urban transit planning of yingtan (2011–2020)," Technical Report, Southeast University, 2011.
- [21] A. G. Karlsruhe, *VISSIM 4.30 User Manual*. PTV, Karlsruhe, Germany, 2007.
- [22] A. Berkhout and J. Righolt, *Chinese Driving Behavior in Nanjing Calibrating VISSIM Parameters*, 2011, <http://www.project-nanjing2009.nl>.

## Research Article

# The Analysis of Braess' Paradox and Robustness Based on Dynamic Traffic Assignment Models

Bai-Bai Fu,<sup>1</sup> Chun-Xue Zhao,<sup>2</sup> and Shu-bin Li<sup>3</sup>

<sup>1</sup> School of Architecture and Urban Planning, Shandong Jianzhu University, Jinan 250101, China

<sup>2</sup> School of Mathematics and Statistics, Anyang Normal University, Anyang 455000, China

<sup>3</sup> Public Security Department, Shandong Police College, Jinan 250014, China

Correspondence should be addressed to Bai-Bai Fu; [fubaibai@163.com](mailto:fubaibai@163.com)

Received 25 July 2013; Accepted 24 October 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Bai-Bai Fu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The investigation of the paradox and robustness about the traffic network is an important branch of the traffic assignment. In this paper, Braess' paradox and robustness of the dynamic traffic network are analyzed by the dynamic traffic assignment models. In addition, the relationship of total costs with different traffic assignment models is discussed. The results show that the paradox only occurs in certain range; the robustness of the network and the relationship of total traffic costs are changed as the traffic demand changes, which provides theoretical guidance for the urban transportation planning.

## 1. Introduction

In recent years, the proposal of the intelligent transportation system (ITS) forces people to investigate the dynamic traffic assignment (DTA). The mathematical programming [1, 2], the optimal control [3, 4], and the variational inequality (VI) [5, 6] were proposed, respectively, and labeled as three main analytical approaches. Specially, the optimal control includes dynamic user equilibrium (DUE) [7–11] and dynamic system optimal (DSO) [12].

Braess' paradox is an important phenomenon during the traffic assignment [13], which has been widely investigated in the scientific literatures [14–16], such as X. Zhang and H. M. Zhang [17] studied the route choices and a novel paradox in queuing networks; Pas and Principio [18] gained the specific range that the paradox occurs when the link cost only depends on the link flow. Braess' paradox is rooted in the essence of the user equilibrium (UE) assignment where each user minimizes his/her own travel time between an origin and a destination. The most previous work mainly focuses on the paradox under the static traffic. In recent years, the paradox of the dynamic traffic assignment began to develop, such that Arnott et al. [19] discussed the properties of the dynamic traffic equilibrium including a paradox; Hallefjord et al. [20]

analyzed the traffic paradox when the travel demand is elastic; Zhang et al. [21] investigated Braess' paradox in the dynamic traffic assignment. In this paper, we investigate the paradox phenomenon of the dynamic traffic network using VI method and assuming the link cost at time  $t$  is related to the flow of this link and the adjacent links at the same paths at time  $t$ .

The robustness, as an important index valuing the stability of the system, is usually used to study the network under the partial degradation. In the process of studying the robustness, it is generally classified into the following two classes according to whether the system structure remains intact: (i) certain component of the network is removed; (ii) no component is removed, but the function of certain portion of the network falls into the perturbations. Based on different classes, different network function measures are proposed to examine the robustness of the network. In this paper, we use the relative total cost index to measure the robustness of the dynamic network considering the influence of the flow on other links when certain component (node or link) is removed.

The paper is organized as follows. We first briefly recall main analytical models of the traffic assignment; then we discuss whether the paradox occurs or not under the given link cost and investigate the relationship between the total

cost under DUE and that under DSO; subsequently, we investigate the robustness of the dynamic network; at last, the conclusions about the results of the paper are given.

## 2. Main Analytical Models

*2.1. The DUE.* In this section, we briefly reviewed dynamic equilibrium models [8–11] of the traffic network. Assume that the traffic network includes a set  $W$  of origin/destination (O/D) pairs with  $n_W$  elements and a set  $P$  of paths joining O/D pairs; DUE is based on the behavioral assumption that only the minimum cost routes are used at each time  $t$ , whose mathematical expression is given as follows:

$$\begin{aligned} C_p(x^*(t)) &= \lambda_w(t), \quad \text{if } x_p^*(t) > 0, \\ C_p(x^*(t)) &\geq \lambda_w(t), \quad \text{if } x_p^*(t) = 0, \end{aligned} \quad (1)$$

where  $C_p(\cdot)$  is the cost on the path  $p$  under the DUO,  $\lambda_w(t)$  is the minimal path costs at time  $t$ ,  $x^*(t)$  is a path flow pattern satisfying the flow conservation at time  $t$ , and  $[0, T]$  denotes the time interval under consideration.

Under the equilibrium state, the link flows and the route flows satisfy the following conservation of flow equations:

$$f_a(t) = \sum_{p \in P} x_p(t) \delta_{ap}, \quad \forall a \in L, \quad (2)$$

where  $f_a(t)$  is the flow on the link  $a$ ,  $x_p(t)$  is the flow on the route  $p$ ,  $\delta_{ap} = 1$  if link  $a$  is contained in route  $p$ , and  $\delta_{ap} = 0$ , otherwise. The traffic demand at time  $t$  must satisfy the following conservation of flow:

$$d_w(t) = \sum_{p \in P_w} x_p(t), \quad \forall w \in W. \quad (3)$$

In addition, the model meets the following nonnegative constraint and boundary initial condition:

$$x_a(t) \geq 0, \quad x_a(0) = 0. \quad (4)$$

*2.2. The DSO.* DSO calls for the flow pattern that minimizes the total travel time in the system at each time  $t$ , whose mathematical expression is shown as follows:

$$\begin{aligned} C'_p(x^*(t)) &= \mu_w(t), \quad \text{if } x_p^*(t) > 0, \\ C'_p(x^*(t)) &\geq \mu_w(t), \quad \text{if } x_p^*(t) = 0, \end{aligned} \quad (5)$$

where  $C'_p(\cdot)$  is the cost on the path  $p$  under the DSO and  $\mu_w(t)$  is the minimal marginal path costs at time  $t$ . The constraint under the DSO is the same as that under DUE and is omitted here.

In addition, the link costs under DUE and DSO satisfy the relationship as follows:

$$c'_a(f_a(t)) = c_a(f_a(t)) + f_a(t) \frac{dc_a(f_a(t))}{df_a(t)}, \quad (6)$$

where  $c_a(\cdot)$ ,  $c'_a(\cdot)$  are costs on link  $a$  under DUE and DSO, respectively; the link cost under the DSO is also known as the marginal link cost.

*2.3. The Variational Inequalities Equivalent to DUE.* Compared with the mathematical programming and optimal control, the variational inequality is the more effective way to describe the dynamic traffic assignment problem. The extensive investigation has gone ahead, such that Frieza and Smith proposed the variational inequality of the dynamic assignment program based on the path cost; Ran and Boyce proposed the variational inequality of the dynamic user optimal model based on the link cost. DUO based on the path cost is equivalent to the variational inequality as follows:

$$\int_0^T \sum_{rs} \sum_p \lambda_p^{rs*}(t) [f_p^{rs}(t) - f_p^{rs*}(t)] dt \geq 0, \quad (7)$$

where  $\lambda_p^{rs*}(t)$  is the path cost on O/D pair  $rs$  at time  $t$  satisfying DUO,  $f_p^{rs}(t)$  is the path flow on O/D pair  $rs$  at time  $t$ , and  $f_p^{rs*}(t)$  is the path flow on O/D pair  $rs$  at time  $t$  satisfying DUE.

## 3. Application of Analytical Models

### 3.1. The Analysis about Paradox

*3.1.1. The Equilibrium Solution of the Four-Link Network under DUE.* As we know, the paradox phenomenon exists in many networks with different topologies; Braess' network has been investigated by a large number of researchers; it is representative because it can simplify the problem and clearly explain traffic phenomena. In this paper, we also consider Braess' network with a single origin  $o$  and a single destination  $r$  in this paper. Figure 1 depicts the network before the addition of the new link  $pq$ . Let the total demand for the travel from origin  $o$  to destination  $r$  be  $d_w(t)$  and  $d_w(t) = t$ ; further, assume that the link cost is a function of the flow not only on this link but also on other links. Because the influence of the adjacent links is dominant, we assume that the other links are only limited to the adjacent links on the same path in this work. At the same time, assume that the problem is symmetric; further, the travel time is link-flow dependent and it increases rapidly as a function of the flow on the link.

Specifically, the link cost functions of the four-link network in Figure 1 are

$$\begin{aligned} c_{op}(t) &= 10(\gamma f_{op}(t) + f_{pr}(t)), \\ c_{qr}(t) &= 10(\gamma f_{qr}(t) + f_{oq}(t)), \\ c_{oq}(t) &= (\gamma f_{oq}(t) + f_{qr}(t)) + 50, \\ c_{pr}(t) &= (\gamma f_{pr}(t) + f_{op}(t)) + 50, \end{aligned} \quad (8)$$

where  $c_{ij}(t)$  is the travel cost on link  $ij$  at time  $t$ ,  $f_{ij}(t)$  is the flow on link  $ij$  at time  $t$ ,  $\gamma$  is the scaling parameter which

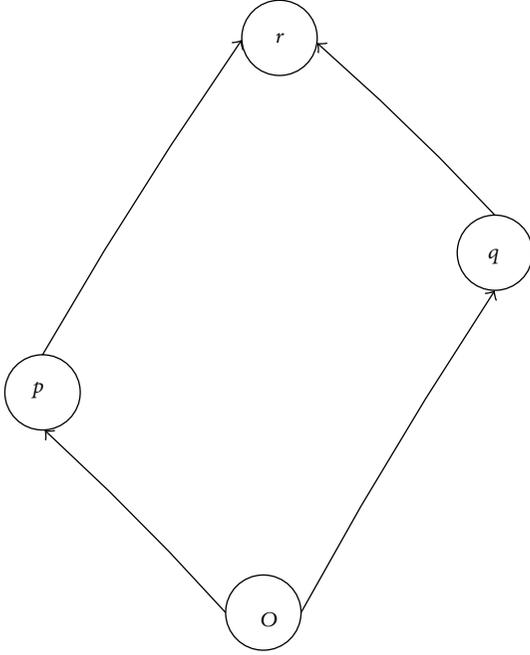


FIGURE 1: The four-link network.

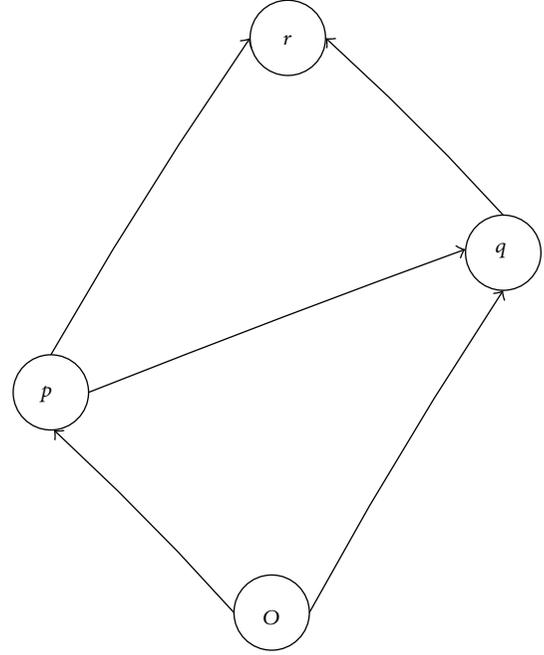


FIGURE 2: The five-link network.

differentiates the influence between the given link only and others, generally,  $\gamma \geq 1$ . In the four-link network, there are two paths from the origin  $o$  to the destination  $r$  and the costs along the paths are given as follows:

$$\begin{aligned} C_1(t) &= c_{op}(t) + c_{pr}(t) = 11(\gamma + 1)f_1(t) + 50, \\ C_2(t) &= c_{oq}(t) + c_{qr}(t) = 11(\gamma + 1)f_2(t) + 50, \end{aligned} \quad (9)$$

where  $C_k(t)$  is the travel cost from  $o$  to  $r$  along path  $k$  at time  $t$ ; in addition,

$$\begin{aligned} f_1(t) &= f_{op}(t) = f_{pr}(t), \\ f_2(t) &= f_{oq}(t) = f_{qr}(t), \end{aligned} \quad (10)$$

where  $f_k(t)$  is the flow from  $o$  to  $r$  along path  $k$  at time  $t$ , the total demand satisfies the following conservation of the flow:  $d_w(t) = f_1(t) + f_2(t)$ . According to the definition of the UE, it is easy to get

$$\begin{aligned} f_1^*(t) &= f_2^*(t) = \frac{t}{2}, \\ C_1(t) &= C_2(t) = \frac{11(\gamma + 1)t}{2} + 50, \end{aligned} \quad (11)$$

where  $f_k^*(t)$  is the equilibrium flow from  $o$  to  $r$  along path  $k$  at time  $t$ . Hence, the total system travel time under DUE before

the addition of the new link is given as follows:

$$C^4(t) = \frac{11(\gamma + 1)t^2}{2} + 50t, \quad (12)$$

where  $C^4(t)$  is the total system travel cost under DUE for the four-link network.

*3.1.2. The Equilibrium Solution of the Five-Link Network under DUE.* After adding the link  $pq$ , there appears a new path  $opqr$  from  $o$  to  $r$  in Figure 2.

The link cost functions are given as follows:

$$\begin{aligned} c_{op}(t) &= 10(\gamma f_{op}(t) + f_{pr}(t) + f_{pq}(t)), \\ c_{qr}(t) &= 10(\gamma f_{qr}(t) + f_{oq}(t) + f_{pq}(t)), \\ c_{oq}(t) &= (\gamma f_{oq}(t) + f_{qr}(t)) + 50, \\ c_{pr}(t) &= (\gamma f_{pr}(t) + f_{op}(t)) + 50, \\ c_{pq}(t) &= (\gamma f_{pq}(t) + f_{op}(t) + f_{qr}(t)) + 10. \end{aligned} \quad (13)$$

The traffic flow on each link is as follows:

$$\begin{aligned} f_{op}(t) &= f_1(t) + f_3(t), \\ f_{pr}(t) &= f_1(t), \\ f_{oq}(t) &= f_2(t), \\ f_{qr}(t) &= f_2(t) + f_3(t), \\ f_{pq}(t) &= f_3(t). \end{aligned} \quad (14)$$

The costs on the paths are

$$C_1(t) = 11(\gamma + 1)f_1(t) + (10\gamma + 11)f_3(t) + 50,$$

$$C_2(t) = 11(\gamma + 1)f_2(t) + (10\gamma + 11)f_3(t) + 50,$$

$$C_3(t) = (10\gamma + 1)(f_1(t) + f_2(t)) + (21\gamma + 22)f_3(t) + 10. \quad (15)$$

The total demand satisfies the conservation of the flow as follows:  $d_w(t) = f_1(t) + f_2(t) + f_3(t)$ .

The variational inequality of the five-link dynamic traffic network over  $t \in [0, T]$  is given as follows:

$$\begin{aligned} & \int_0^T (11(\gamma + 1)f_1^*(t) + (10\gamma + 11)f_3^*(t) + 50) \\ & \times (f_1(t) - f_1^*(t)) \\ & + (11(\gamma + 1)f_2^*(t) + (10\gamma + 11)f_3^*(t) + 50) \\ & \times (f_2(t) - f_2^*(t)) \\ & + ((10\gamma + 1)(f_1^*(t) + f_2^*(t)) \\ & + (21\gamma + 22)f_3^*(t) + 10)(f_3(t) - f_3^*(t)) dt \geq 0. \end{aligned} \quad (16)$$

Because

$$\begin{aligned} d_w(t) &= f_1(t) + f_2(t) + f_3(t), \\ d_w(t) &= f_1^*(t) + f_2^*(t) + f_3^*(t), \\ f_1^*(t) &= f_2^*(t), \end{aligned} \quad (17)$$

we have

$$\begin{aligned} & \int_0^T ((13\gamma + 11)f_1^*(t) - 11(\gamma + 1)t + 40) \\ & \times (f_1(t) + f_2(t) - 2f_1^*(t)) dt \geq 0; \end{aligned} \quad (18)$$

if  $f_1^*(t) = 0$ , and we have  $-11(\gamma + 1)t + 40 \geq 0$ , then  $t \leq 40/11(\gamma + 1)$ ; if  $f_3^*(t) = 0$ , and at this time,  $f_1^*(t) = t/2$ , then  $t > 80/(9\gamma + 11)$ ; when  $t \in (40/11(\gamma + 1), 80/(9\gamma + 11))$ ,

$$\begin{aligned} f_1^*(t) = f_2^*(t) &= \frac{11(\gamma + 1)t - 40}{13\gamma + 11}, \\ f_3^*(t) &= \frac{-(9\gamma + 11)t + 80}{13\gamma + 11}. \end{aligned} \quad (19)$$

Assume  $\gamma = 2$ ; the equilibrium flow of the five-link network is pictured in Figure 3, in ranges I and II  $[0, 33/40]$ , only the new path is used; in range III  $(33/40, 80/29)$ , all three are used; in range IV  $[80/29, +\infty)$ , only the first two paths are used; that is, the third path is never used when  $t > 80/29$ .

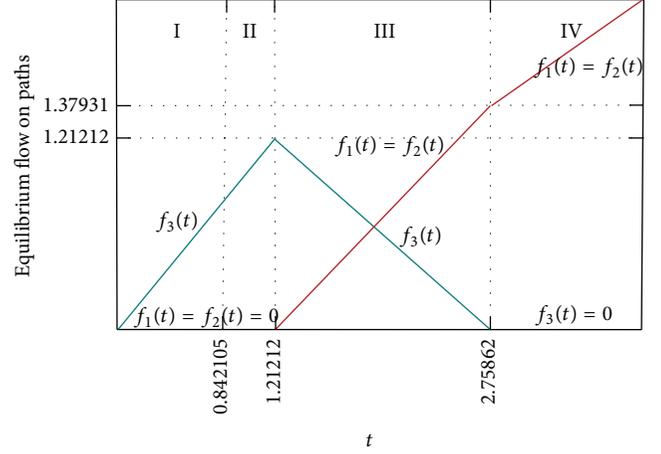


FIGURE 3: The equilibrium flow of the dynamic network.

Corresponding to different ranges, the total travel cost of five-link network is given as follows:

$$\begin{aligned} C^5(t) &= \begin{cases} (21\gamma + 22)t^2 + 10t, & \text{if } t \leq \frac{40}{11(\gamma + 1)}, \\ t \left[ \frac{(31\gamma^2 + 33\gamma)t + 40(9\gamma + 11)}{13\gamma + 11} + 50 \right], & \text{if } \frac{40}{11(\gamma + 1)} < t < \frac{80}{9\gamma + 11}, \\ \frac{11(\gamma + 1)t^2}{2} + 50t, & \text{if } t \geq \frac{80}{9\gamma + 11}, \end{cases} \end{aligned} \quad (20)$$

where  $C^5(t)$  is the total system travel cost under DUE for the five-link network.

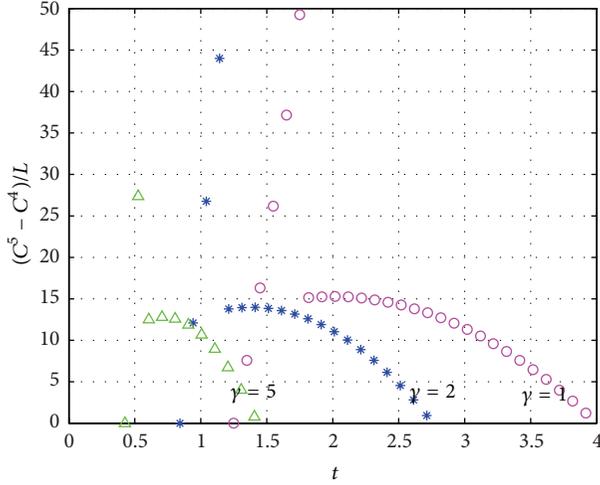
**3.1.3. The Paradox of Dynamic Traffic Networks.** Letting  $C^5(t) > C^4(t)$ , we get  $t \in (80/(31\gamma + 33), 80/(9\gamma + 11))$ ; letting  $\gamma = 2$ , the paradox occurs in ranges II and III in Figure 3.

In order to capture the trend of the paradox when we take different values of  $\gamma$ , we give the definition of the average difference of two functions  $F_1, F_2$  in range  $[a, b]$  ( $[a, b][a, b], (a, b)$ ),  $a > b$  as follows:

$$D(F_1, F_2) = \frac{|F_1 - F_2|}{b - a}. \quad (21)$$

We assume that the severity of the paradox is proposal to the average difference and use of the average difference of  $C^4, C^5$  in range  $(80/(31\gamma + 33), 80/(9\gamma + 11))$  to represent the severity of the paradox.

In Figure 4, we discuss the situations when  $\gamma = 1, 2, 5$ , respectively, and find that the areas where the paradox occurs are different when the values of  $\gamma$  are different; thus we may take some measures to control the influence between the links


 FIGURE 4: The severity that the paradox occurs when  $\gamma$  is different.

as time and traffic demand change in order to avoid or decrease the occurrence of the paradox; in addition, in most range, the greater  $\gamma$  is, the smaller the range of the traffic demand in which the paradox occurs is; that is, the severity and the range in which the paradox occurs increase with the influence of other links decreasing, which reminds us to let down the influence between links as much as possible if we wish to decrease the occurrence of the paradox.

**3.1.4. The Effect of the Adding Link under DSO.** As known, when the given link cost is a function of the flow on the given link only, adding a new link did not reduce the total system travel time even under SO [18]. When the link cost does not only depend on the flow on the given link only, we discuss whether the phenomenon occurs or not in the following. As we know, DSO is obtained by charging users the marginal cost of traveling (see formula (6)); for the link cost in this work, the corresponding path marginal cost equations are given as follows:

$$\begin{aligned}
 c'_{op}(t) &= 10(2\gamma f_{op}(t) + f_{pr}(t) + f_{pq}(t)), \\
 c'_{qr}(t) &= 10(2\gamma f_{qr}(t) + f_{oq}(t) + f_{pq}(t)), \\
 c'_{oq}(t) &= (2\gamma f_{oq}(t) + f_{qr}(t)) + 50, \\
 c'_{pr}(t) &= (2\gamma f_{pr}(t) + f_{op}(t)) + 50, \\
 c'_{pq}(t) &= (2\gamma f_{pq}(t) + f_{op}(t) + f_{qr}(t)) + 10.
 \end{aligned} \tag{22}$$

The cost on the paths is as follows:

$$\begin{aligned}
 C'_1(t) &= 11(2\gamma + 1)f_1(t) + (20\gamma + 11)f_3(t) + 50, \\
 C'_2(t) &= 11(2\gamma + 1)f_2(t) + (20\gamma + 11)f_3(t) + 50, \\
 C'_3(t) &= (20\gamma + 11)(f_1(t) + f_2(t)) \\
 &\quad + 2(21\gamma + 11)f_3(t) + 10.
 \end{aligned} \tag{23}$$

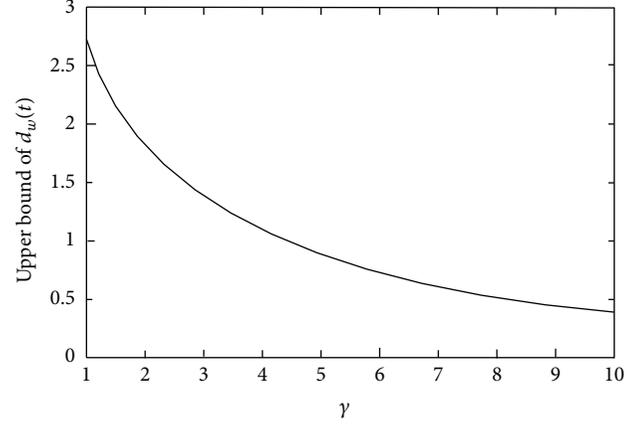


FIGURE 5: The bound above which an adding link cannot make the total costs increase under DSO.

Then the variational inequality of the five-link dynamic traffic network over  $t \in [0, T]$  under DSO is given as follows:

$$\begin{aligned}
 &\int_0^T (11(2\gamma + 1)f_1^*(t) + (20\gamma + 11)f_3^*(t) + 50) \\
 &\quad \times (f_1(t) - f_1^*(t)) \\
 &\quad + (11(2\gamma + 1)f_2^*(t) + (20\gamma + 11)f_3^*(t) + 50) \\
 &\quad \times (f_2(t) - f_2^*(t)) \\
 &\quad + ((20\gamma + 11)(f_1^*(t) + f_2^*(t)) \\
 &\quad \quad + 2(21\gamma + 11)f_3^*(t) + 10) \\
 &\quad \times (f_3(t) - f_3^*(t)) dt \geq 0.
 \end{aligned} \tag{24}$$

Because of formula (17), the equation can be given as follows:

$$\begin{aligned}
 &\int_0^T ((26\gamma + 11)f_1^*(t) - 11(2\gamma + 1)t + 40) \\
 &\quad \times (f_1(t) + f_2(t) - 2f_1^*(t)) dt \geq 0.
 \end{aligned} \tag{25}$$

Let  $f_3^*(t) = 0$ ; at this time,  $f_1^*(t) = f_2^*(t) = t/2$  and  $(26\gamma + 11)(t/2) - 11(2\gamma + 1)t + 40 \geq 0$ ; we have  $t \leq 80/(18\gamma + 11)$ ; when  $t \in [0, 80/(18\gamma + 11)]$ , the adding link makes sense under DSO.

In the following, we give the trend of the upper bound under which the adding link works under DSO as the parameter  $\gamma$  changes in Figure 5 and find that the bound becomes smaller as  $\gamma$  increases, which explains that the less the influence of the other links is, the less the possibility that the adding link works under DSO, which warns us of improving the influence between the links appropriately if we want to make the adding link work under DSO.

**3.1.5. Relationship between the Total Cost under DUE and DSO.** There is much research about the relationship of the total costs under different assignments, in which [22] we know that the solution under DUE and that under DSO are

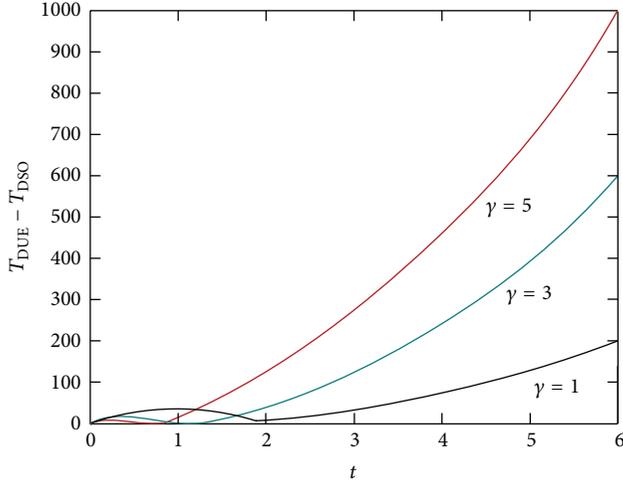


FIGURE 6: Trend of  $T_{DUE} - T_{DSO}$  with  $t$  changing when the values of  $\gamma$  are different.

close under the free flow state, and the difference between both assignments is great when the traffic is congested in the kind of networks. In this section, we discuss the situation of the flow of the link and the flow of other links. Define  $T_{DUE} - T_{DSO}$  as the distance of the total costs between DUE and DSO. In this paper, we capture the gap of the total costs between the two assignments using the distance of the total costs between DUE and DSO. In Figure 6, we give the trend of the gap when  $\gamma$  is 1, 3, and 5, respectively and find that the gap increases with the time being over and demand accumulating; in addition, the gap increases with  $\gamma$  increasing in the area of the great demand. It shows, with  $\gamma$  increasing, that the influence of other links decreases; that is, it is reverse with the idea of DSO ignoring the influence of other links; thus the gap increases.

**3.2. Robustness of the Dynamic Network.** The network robustness has been widely investigated in many literatures, in which the importance identification of the network components is an important branch [23, 24]. In this section, we use the relative total cost index to determine the importance of the components of the dynamic network.

The total cost on the link  $a$  is denoted as follows:

$$\hat{c}_{ij}(t) = c_{ij}(t) f_{ij}(t), \quad (26)$$

where  $\hat{c}_{ij}(t)$  is the total cost on the link  $ij$  at time  $t$ ,  $c_{ij}(t)$  is the unit cost on the link  $ij$  at the time  $t$ , and  $f_{ij}(t)$  is the flow on the link  $ij$  at time  $t$ . Hence, the total cost  $TC(t)$  on a network at time  $t$  is given by

$$TC(t) = \sum_{ij \in L} \hat{c}_{ij}(t), \quad (27)$$

where  $L$  is the set of the links. The link flows  $f(t)$  satisfy the nonnegativity and conservation of the flow conditions. Then we have the relative total cost index as follows:

$$\Gamma^l(t) = \frac{TC_{G-l}(t) - TC(t)}{TC(t)}, \quad (28)$$

where  $\Gamma^l(t)$  denotes the important indicator for the link  $l$  under the DUE at time  $t$  and  $TC_{G-l}(t)$  denotes the total cost generated under DUE when the link  $l$  is removed from the network at time  $t$ . Similarly, the importance indicator for the node  $M$  is denoted as follows:

$$\Gamma^M(t) = \frac{TC_{G-M}(t) - TC(t)}{TC(t)}, \quad (29)$$

where  $\Gamma^M(t)$  is the importance indicator for the node  $M$  under DUE at time  $t$ ,  $TC_{G-M}(t)$  is the total cost when the node  $M$  is removed from the network under DUE at time  $t$ .

In the following, we use the network in Figure 2 and assume  $\gamma = 2$  to investigate the importance of components of the dynamic network. During the process of the calculation, it can be found that the different traffic demands correspond to different traffic assignments with different importance components; the range of  $t$  is divided into four different ranges as follows.

Demand Range I:

$$t \in \left[0, \frac{16}{19}\right). \quad (30)$$

We know that, in this demand range, only the third path is used at the equilibrium; hence, we have the equilibrium path flow pattern  $f(t)_1 = f_2(t) = 0, f_3(t) = t$ , and the total cost is  $TC(t) = 64t^2 + 10t$ .

Demand Range II:

$$t \in \left[\frac{16}{19}, \frac{40}{33}\right). \quad (31)$$

We know that, in this demand range, only the third path is used at the equilibrium, hence; we have the equilibrium path flow pattern  $f(t)_1 = f_2(t) = 0, f_3(t) = t$ , and the total cost is  $TC(t) = 64t^2 + 10t$ .

Demand Range III:

$$t \in \left[\frac{40}{33}, \frac{80}{29}\right). \quad (32)$$

We know that, in this range of the demand, all three paths are used at the equilibrium; we now have that  $f_1(t) = f_2(t) = (33t - 40)/37, f_3(t) = (80 - 29t)/37$ , and the total cost is  $TC(t) = (190t^2 + 2640t)/37$ .

Demand Range IV:

$$t \in \left[\frac{80}{29}, +\infty\right). \quad (33)$$

We know that only the first two paths are used at the equilibrium and the Braess' paradox vanishes. Hence, we have the equilibrium path flow pattern  $f_1(t) = f_2(t) = t/2, f_3(t) = 0$ , and the total cost is  $TC(t) = (33t^2/2) + 50t$ .

The ranking of the component is the sequence of its important value. Then the important values and rankings of different components in different ranges are given in Tables 1, 2, 3, and 4.

From Tables 1–4, the important values and the rankings of the different components are different; at the same time,

TABLE 1: Important values of links in different ranges.

	I	II	III	IV
$op$	$\frac{40 - 31t}{64t + 10}$	$\frac{40 - 31t}{64t + 10}$	$\frac{1031t - 1160}{190t + 3010}$	$\frac{33t}{33t + 100}$
$oq$	0	0	$\frac{35937t - 43560}{6650t + 105350}$	$\frac{1147t - 160}{1155t + 3500}$
$pr$	0	0	$\frac{35937t - 43560}{6650t + 105350}$	$\frac{1147t - 160}{1155t + 3500}$
$qr$	$\frac{40 - 31t}{64t + 10}$	$\frac{40 - 31t}{64t + 10}$	$\frac{1031t - 1160}{190t + 3010}$	$\frac{33t}{33t + 100}$
$pq$	$\frac{5(18 - 19t)}{4(32t + 5)}$	$\frac{5(18 - 19t)}{4(32t + 5)}$	$\frac{841t - 2320}{380t + 6020}$	0

TABLE 2: Important rankings of links in different ranges.

	I	II	III	IV
$op$	1	1	1	1
$oq$	3	2	2	2
$pr$	3	2	2	2
$qr$	1	1	1	1
$pq$	2	3	3	3

TABLE 3: Important values of nodes in different ranges.

	I	II	III	IV
$o$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
$p$	$\frac{40 - 31t}{64t + 10}$	$\frac{40 - 31t}{64t + 10}$	$\frac{1031t - 1160}{190t + 3010}$	$\frac{33t}{33t + 100}$
$q$	$\frac{40 - 31t}{64t + 10}$	$\frac{40 - 31t}{64t + 10}$	$\frac{1031t - 1160}{190t + 3010}$	$\frac{33t}{33t + 100}$
$r$	$+\infty$	$+\infty$	$+\infty$	$+\infty$

TABLE 4: Important rankings of nodes in different ranges.

	I	II	III	IV
$o$	1	1	1	1
$p$	2	2	2	2
$q$	2	2	2	2
$r$	1	1	1	1

the important rankings of the component are different in different ranges of the time, which show the robustness of the network changes when different components collapse.

Owing to considering the influence of other links and using different efficiency measures, we find that the important values of the corresponding nodes and links are different from the results of Nagurny and Qiang [25], but the important ranking of the corresponding components is the same as their results; there are two possible reasons as follows: (i) the purpose and essence of the different efficiency measures are consistent; (ii) in this work, we assume  $\gamma = 2$ ; that is, the influence of other links is secondary.

In addition, from Sections 3.1.3 and 3.2, we find that the importance of the adding link is small in the area that Braess'

paradox occurs and the adding link is not used, which shows that the robustness of the network is related to whether the paradox occurs. Based on the preceding study, we consider the influence of other links in this paper, which better provide the guidance of the traffic assignment.

### 4. Conclusion

In this work, we investigate the paradox phenomenon and the robustness of the dynamic traffic network, in which the cost influenced by other link flows is considered. The four-link network and the five-link network are used in the analysis, which reveal that the paradox occurs only in certain range. The difference of the total costs between DUE and DSO changes as the traffic demand changes. The important values and important rankings of network components are changing at different demand levels. The results show that whether the paradox occurs or not and the robustness are both affected by the traffic demand, and the robustness of the network is related to the occurrence of the paradox, which provides a theoretical basis to the traffic assignment. The more complex networks are closer to the reality and needed to be further investigated.

### Acknowledgments

The authors would like to thank the anonymous referees for their useful comments on this paper. This work is supported by the National Natural Science Foundation of China (71171124, 10871219).

### References

- [1] J. R. Birge and J. K. Ho, "Optimal flows in stochastic dynamic networks with congestion," *Operations Research*, vol. 41, no. 1, pp. 203–216, 1993.
- [2] B. N. Janson, "Dynamic traffic assignment for urban road networks," *Transportation Research B*, vol. 25, no. 2-3, pp. 143–161, 1991.
- [3] B.-W. Wie, T. L. Friesz, and R. L. Tobin, "Dynamic user optimal traffic assignment on congested multidestination networks," *Transportation Research B*, vol. 24, no. 6, pp. 431–442, 1990.
- [4] B. Ran, D. E. Boyce, and L. J. LeBlanc, "New class of instantaneous dynamic user-optimal traffic assignment models," *Operations Research*, vol. 41, no. 1, pp. 192–202, 1993.
- [5] H.-K. Chen and C.-F. Hsueh, "A model and an algorithm for the dynamic user-optimal route choice problem," *Transportation Research B*, vol. 32, no. 3, pp. 219–234, 1998.
- [6] R. Jayakrishnan, W. K. Tsai, and A. Chen, "A dynamic traffic assignment model with traffic-flow relationships," *Transportation Research C*, vol. 3, no. 1, pp. 51–72, 1995.
- [7] X. J. Nie, *The study of dynamic user-equilibrium traffic assignment [Ph.D. thesis]*, University of California, Davis, Calif, USA, 2003.
- [8] S. C. Dafermos and F. T. Sparrow, "The traffic assignment problem for a general network," *Journal of Research of the National Bureau of Standards*, vol. 73, pp. 91–118, 1969.
- [9] M. J. Smith, "The existence, uniqueness and stability of traffic equilibria," *Transportation Research B*, vol. 13, no. 4, pp. 295–304, 1979.

- [10] M. Ng and S. T. Waller, "A dynamic route choice model considering uncertain capacities," *Computer-Aided Civil and Infrastructure Engineering*, vol. 27, no. 4, pp. 231–243, 2012.
- [11] Q. Meng and H. L. Khoo, "A computational model for the probit-based dynamic stochastic user optimal traffic assignment problem," *Journal of Advanced Transportation*, vol. 46, no. 1, pp. 80–94, 2012.
- [12] S. Dafermos, "Traffic equilibrium and variational inequalities," *Transportation Science*, vol. 14, no. 1, pp. 42–54, 1980.
- [13] D. Braess, "Über ein paradoxon aus der verkehrsplanung," *Unternehmensforschung*, vol. 12, pp. 258–268, 1968.
- [14] S. Dafermos and A. Nagurney, "On some traffic equilibrium theory paradoxes," *Transportation Research B*, vol. 18, no. 2, pp. 101–110, 1984.
- [15] T. Akamatsu and B. Heydecker, "Detecting dynamic traffic assignment capacity paradoxes in saturated networks," *Transportation Science*, vol. 37, no. 2, pp. 123–138, 2003.
- [16] T. Roughgarden, *Selfish Routing and the Price of Anarchy*, MIT Press, 2005.
- [17] X. Zhang and H. M. Zhang, "Simultaneous departure time/route choices in queueing networks and a novel paradox," *Networks and Spatial Economics*, vol. 10, no. 1, pp. 93–112, 2010.
- [18] E. I. Pas and S. L. Principio, "Braess' paradox: some new insights," *Transportation Research B*, vol. 31, no. 3, pp. 265–276, 1997.
- [19] R. Arnott, A. de Palma, and R. Lindsey, "Properties of dynamic traffic equilibrium involving bottlenecks, including a paradox and metering," *Transportation Science*, vol. 27, no. 2, pp. 148–160, 1993.
- [20] A. Hallefjord, K. Jörnsten, and S. Storøy, "Traffic equilibrium paradoxes when travel demand is elastic," *Asia-Pacific Journal of Operational Research*, vol. 11, no. 1, pp. 41–50, 1994.
- [21] X. Zhang, W. H. K. Lam, and H.-J. Huang, "Braess's paradoxes in dynamic traffic assignment with simultaneous departure time and route choices," *Transportmetrica*, vol. 4, no. 3, pp. 209–225, 2008.
- [22] J. N. Prashker and S. Bekhor, "Some observations on stochastic user equilibrium and system optimum of traffic assignment," *Transportation Research B*, vol. 34, no. 4, pp. 277–291, 2000.
- [23] A. Nagurney and Q. Qiang, "A relative total cost index for the evaluation of transportation network robustness in the presence of degradable links and alternative travel behavior," *International Transactions in Operational Research*, vol. 16, no. 1, pp. 49–67, 2009.
- [24] A. Nagurney, Q. Qiang, and L. S. Nagurney, "Environmental impact assessment of transportation networks with degradable links in an era of climate change," *International Journal of Sustainable Transportation*, vol. 4, no. 3, pp. 154–171, 2010.
- [25] A. Nagurney and Q. Qiang, "A network efficiency measure for congested networks," *Europhysics Letters*, vol. 79, no. 3, Article ID 38005, 5 pages, 2007.

## Research Article

# Development of Degree-of-Priority Based Control Strategy for Emergency Vehicle Preemption Operation

Jiawen Wang, Wanjing Ma, and Xiaoguang Yang

Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China

Correspondence should be addressed to Wanjing Ma; [mawanjing@tongji.edu.cn](mailto:mawanjing@tongji.edu.cn)

Received 22 August 2013; Accepted 5 November 2013

Academic Editor: Huimin Niu

Copyright © 2013 Jiawen Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a degree-of-priority based control strategy for emergency vehicle preemption operation to decrease the impacts of emergency vehicles on normal traffic. The proposed model features its effectiveness to the following three aspects: (1) a multilayer fuzzy model was established to determine the degree-of-priority based on emergency vehicle preemption demand intensity and preemption influence intensity; (2) for emergency vehicles with proper classification, a travel time estimation model for emergency traffic was formulated, an optimal emergency route determines model based on the level of priority of emergency events, and the emergency vehicle travel time was developed to minimize evacuation time as well as minimize the adverse impacts of preemption on normal traffic; and (3) a conditional traffic signals priority control method at each intersection of the evacuation route was built, so that traffic queue at each intersection can be cleared before the arrival of emergency vehicles. A simulation model based on field data was developed, and the performance of the proposed strategy was compared with the conventional local detection based method under the microscopic simulation model. The results validated the efficiency of the proposed strategy in terms of minimizing the delay of emergency vehicles and reducing adverse impacts on normal traffic.

## 1. Introduction

Providing safe and fast driving environment for emergency vehicles to reduce travel time and delay is a critical issue in traffic evacuation. Under effective preemption, ones can reach their destinations at the earliest possible time which is one of most critical factors in saving lives and reducing property loss. At the same time, reducing the adverse impacts of emergency vehicles on normal traffic, so that they can cause the least disturbance to network traffic flow, is the key to avoid the grid-lock caused by emergency accidents [1, 2]. While substantial progress has been made in the areas of vehicle detection and communication technologies to increase the efficiency of emergency vehicles, current state-of-the-art in signal preemption in China has not reached the point where signal clearance strategy, considering the adverse impact on normal traffic, can be automatically generated and implemented in real time.

To date, most preemption systems developed operate on a single-intersection basis and require local detection of an

emergency vehicle to activate a signal preemption sequence at each intersection [3, 4]. Existing signal preemption methods can be classified into several categories such as optical, infrared light, acoustic, special types of loop detection, and GPS-based systems [5, 6]. The optical systems that are developed in the 1960's use a strobe-lamp on the vehicle and an optical sensor per approach to an intersection requiring a clear line-of-sight path between the vehicle and the intersection [7]. The sound-based systems use the directional microphones installed at an intersection to detect the siren of vehicles approaching a given intersection; therefore, no special equipment is required for the emergency vehicles [8]. In a GPS-based system being operated in Taicang, China, both an emergency vehicle and an intersection are equipped with a GPS receiver and a radio transceiver for two-way communication. In recent studies a route-based dynamic strategy was developed for efficient preemption of traffic signals for emergency vehicles in real time by Kwon et al. [9] And Louisell and Collura [10] proposed a simple algorithm to estimate emergency vehicle travel time savings based on

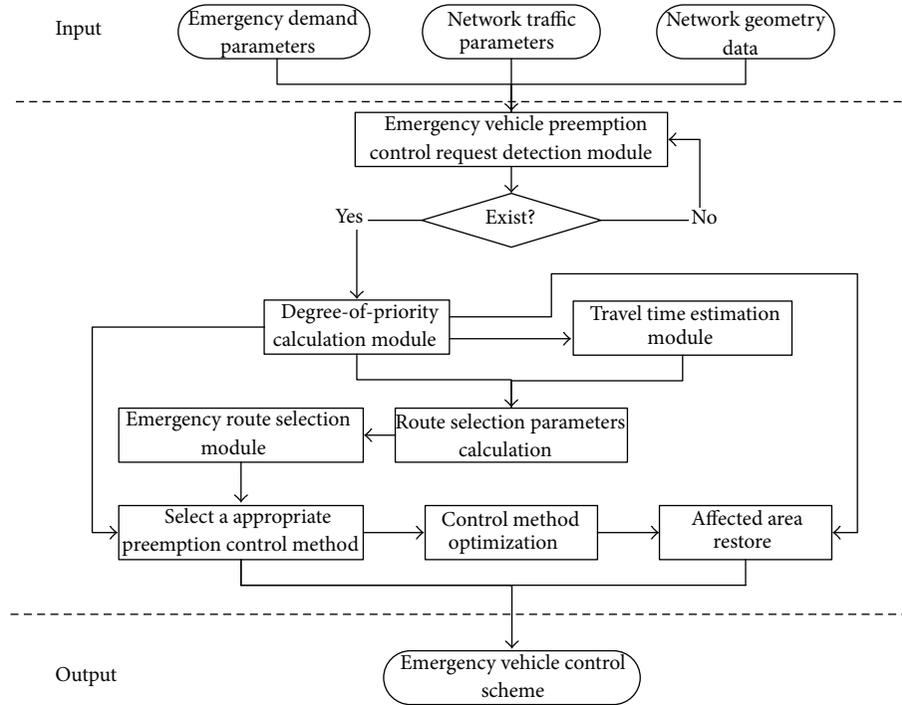


FIGURE 1: Structure of degree-of-priority based signal preemption strategies.

a field operational. Mussa and Selekwia [11] reported on the development of a transition procedure based on the quadratic optimization method that aimed at reducing disutility measures to motorists during the transition period. Haghani et al. [12] concentrated on developing an optimization model for developing flexible dispatching strategies that take the advantage of available real-time travel time information. In recent years, Yun et al. [13, 14] have optimized the exit phase controls for emergency vehicle preemption and compared the emergency vehicle preemption methods with a hardware-in-the-loop simulation. He et al. [15] presented a heuristic algorithm for traffic signal control with simultaneous multiple priority requests at isolated intersections in the context of vehicle-to-infrastructure communications being available on priority vehicles, and the method could reduce average bus delay in congested conditions by about 50%. Savolainen et al. [16] developed a dynamically activated emergency vehicle alert system to provide an additional visual cue to motorists of an impending emergency vehicle's approach. Route selection is one of the fundamental problems in emergency logistics management [17–20]. In order to send the commodities as quickly as possible, the path which costs least should be selected. A vast amount of the literature [21–25] has been produced up until now.

As indicated above, while there has been substantial progress in developing local preemption technologies, and much research has focused on the impact of transit priority [26–28], the impact of EV preemption has not been well studied; a few research studies on decreasing the adverse impacts of preemption on normal traffic have been found in the literature. Qin and Khan [29] report two new control

strategies for emergency vehicle signal preemption which reduce the response time and minimize the impact of EV operation on general traffic. Developing degree-of-priority based control strategy that can provide an efficient and safe traveling environment for emergency vehicles with minimum disruption on network traffic is of critical importance in managing urban traffic.

Based on the above achievements, this paper presents a dynamic preemption approach that combines a degree of control priority classification procedure and a route-based preemption method to provide the most appropriate route and control strategy for an emergency vehicle under a given network, traffic conditions, and emergency issue conditions. The proposed method was evaluated in part of the roads network in Beijing, China, using a microscopic network simulation model, and its effectiveness was compared with the existing local-detection based method.

This paper is organized as follows: the first part is the introduction of emergency vehicle preemption methods. The second part is the model of degree of priority classification and travel time prediction. The third part is development of emergency vehicle preemption strategy and performance evaluation. Last part is conclusion.

## 2. Degree-of-Priority Based Signal Preemption

Figure 1 shows the structure of the degree-of-priority based signal preemption strategy developed in this study. Input data include emergency demand parameters, network traffic parameters, and network data. Emergency demand parameters are obtained when emergency issue happens.

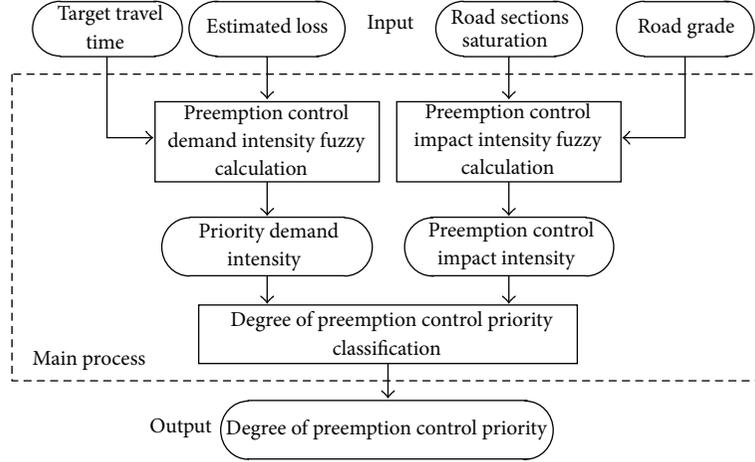


FIGURE 2: Emergency vehicle preemption signal control priority classification process.

Network traffic parameters are continuously collected by field detectors. Network geometry data are static data provided by government. Control strategy for emergency vehicle preemption will be started as soon as emergency issue happens, and the first step is classifying degree-of-priority of emergency vehicles in each road section, which corresponds to intensity needs of emergency vehicle preemption. Then travel time of emergency flow in each road section is predicted based on the input data and degree of priority. Next, two-objective optimization model is established to calculate route selection parameters and select the optimal route for emergency flow. In this research, the well-known Dijkstra algorithm [30] is adopted to find the optimal route that has the minimum route selection parameter for a given origin-destination pair. At last, signal control method is developed to optimize the emergency vehicle preemption plan, and affected area restore control plan is proposed.

### 3. Degree of Priority Classification Model

Degree of priority is an important parameter of emergency vehicle in this research, which corresponds to the important level and degree of influence of emergency vehicle in each road section. Because both importance level of emergency vehicle and degree of influence to normal network traffic flow are difficult to measure accurately, degree of priority could be calculated by the genetic algorithm or fuzzy algorithm. In this study, researchers developed a three-step, multilayer fuzzy algorithm to calculate the degree of priority in three steps.

To calculate preemption demand intensity, the seriousness and urgency of the emergency issue should be estimated based on desired travel time of emergency vehicle and estimated loss of the emergency issue. In the fuzzy reasoning of preemption demand intensity calculation, fuzzy inputs are the estimated loss parameter  $P_{lk}$  of emergency issue  $k$  and the urgency parameter  $U_k$ . Fuzzy output is preemption demand intensity  $I_{pd}$ . For the convenience of computing, in this research the ranges of estimated loss parameter  $P_{lk}$  and the urgency parameter  $U_k$  are defined as 0 to 1 (see Figure 2).

The estimated loss parameter  $P_{lk}$  can be calculated as follows:

$$P_{lk} = \frac{\text{Log}(L_k + nL_k^{(P)})}{10}. \quad (1)$$

In (1),  $P_{lk}$  is the estimated loss parameter of emergency issue  $k$ .  $L_k$  is the estimated economic loss of emergency issue  $k$  with 10,000 CNY as its unit.  $L_k^{(P)}$  is the conversion coefficient of casualties.  $n$  is the estimated number of casualties. According to this formula, emergency events anticipated to have a higher estimated loss would result in a higher intensity of response.

The urgency parameter  $U_k$  can be calculated as follows:

$$U_k = 1 - \frac{\sum_{n=1}^j (L_{kn}/v_{mn})}{T_t}. \quad (2)$$

In (2),  $U_k$  is the urgency parameter of emergency issue  $k$ .  $T_t$  is the desired travel time of emergency vehicle appointed by decision makers.  $L_{kn}$  is the length of road section  $n$  in the shortest route  $j$ .  $v_{mn}$  is the top speed of the emergency vehicle in road section  $n$ . From the formula we can find with the decrease of the urgency parameter  $U_k$ , the emergency vehicle is more urgent. It makes these two parameters fuzzy by membership function, as shown in Figures 3(a) and 3(b), to calculate the preemption demand intensity.

Fuzzy rules of preemption demand intensity classification are shown in Table 1(a). In Table 1, {VH, H, M, L, VL} is short for {Very High, High, Medium, Low, Very Low}.

The membership function of preemption demand intensity is triangular, and the range is [0, 1]. Then, the exact value function of preemption demand intensity is obtained through the center-of-gravity defuzzification.

To consider the impact of emergency vehicle preemption on normal network traffic flow, preemption impact intensity is proposed to reflect the degree of this impact. Fuzzy inputs of preemption impact intensity fuzzy calculation are city road grade  $v_{dj}$  of road section  $j$  and saturation  $x_j$  of road section  $j$ . The city road grade is divided into expressway, arterial road,

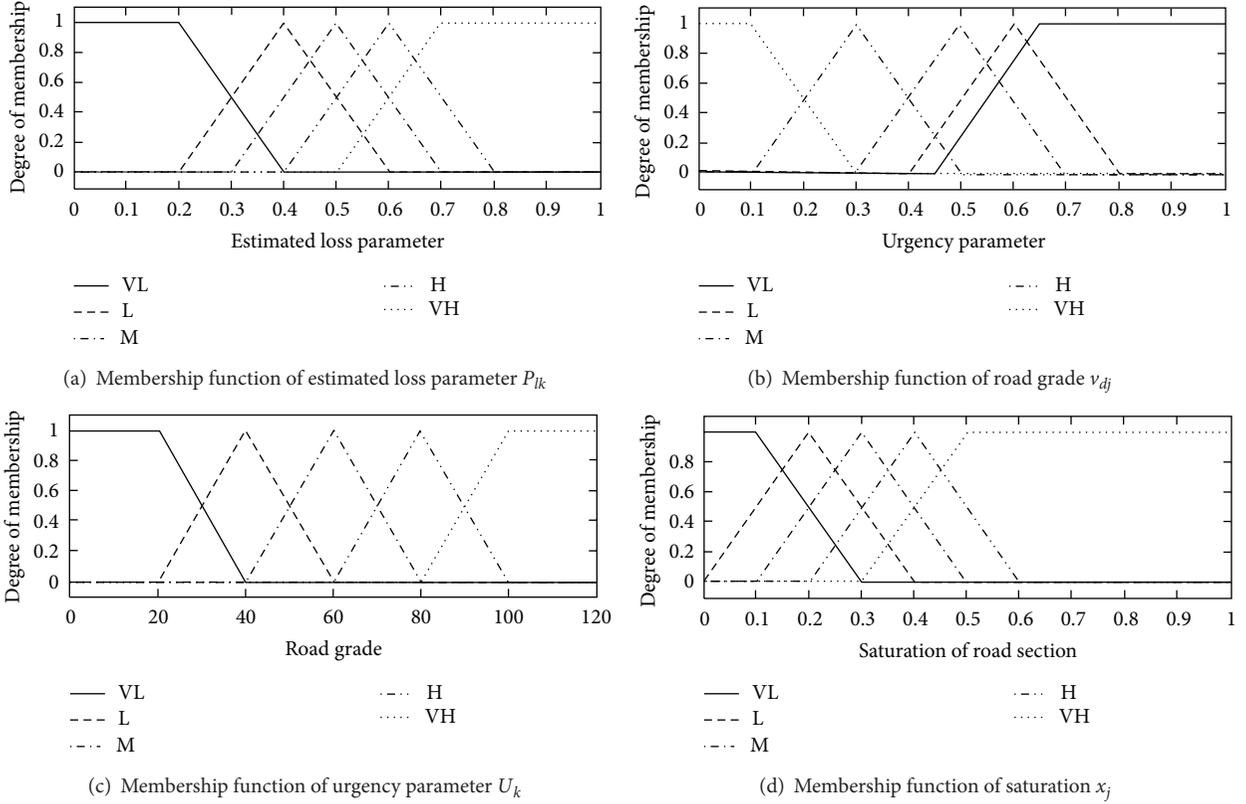


FIGURE 3: Membership function in degree of priority classification model.

secondary trunk road, and branch road, which are indicated by value of  $v_{dj}$ . {80, 60, 40, 30} indicates {Expressway, Arterial Road, Secondary Trunk Road, Branch Road}, respectively. Saturation of road section is calculated by volumes and capacity of the road section. It makes these two parameters fuzzy by membership function that is shown in Figures 3(c) and 3(d) to calculate preemption impact intensity. Fuzzy rules of preemption impact intensity calculation are shown in Table 1(b).

The membership function of preemption impact intensity is triangular, and the range is  $[0, 1]$ . Then, the exact value function of preemption impact intensity is obtained through the center-of-gravity defuzzification.

To calculate degree of priority of emergency vehicle on each road section, fuzzy rules of emergency vehicle preemption priority classification are used. The fuzzy rule is shown in Table 1(c).

In Table 1(c), when preemption impact intensity is very high, the degree of priority is classified as a low value to avoid the impact on normal network traffic flow. And when preemption impact intensity is very low, normal network traffic flow volume is low. So the road impedance to emergency vehicle is low enough to classify degree of priority as a low value. Therefore, when Preemption Impact Intensity is medium, a high value is given according to these rules.

The membership function of emergency vehicle degree of priority is triangular, and the range is  $[0, 1]$ . Then, the exact value function of emergency vehicle preemption priority is

obtained through the center-of-gravity defuzzification. The emergency vehicle degree of priority  $D_p$  is an important input in selecting control method.

#### 4. Travel Time Prediction Model for Emergency Vehicle

Road section clearance time,  $\Delta t$ , is an important parameter that affects total travel time and delay of emergency vehicles. It stands for the time difference between the moment  $t_e$  when the sections are changed into emergency state and the moment  $t_{in}$  when the emergency vehicles enter the sections. The road section clearance time of different control methods is not the same, and it is related to emergency vehicle degree of priority of the road section. The higher degree of priority, the longer road section clearance time. In the urban road system, the difference of traffic characteristics vehicles will become discrete. The traffic flow setup from the upper intersection will become discrete, and the length of platoon will become long before the platoon arrives at the lower intersection. This phenomenon is called platoon dispersion. Previous studies have demonstrated that the speeds of vehicles follow Gaussian distributions.

In the multilane sections, lane-clear speed varies directly with mean time-headway. Lane-clear speed has the dimensions of pcu/km·s.

The lane-change behavior is very fast then the distance of platoon moving in the lane-change period is short,

TABLE 1: Fuzzy rules in degree of priority classification model.

(a) Fuzzy rules of preemption demand intensity calculation					
$U_k$	$I_{pd}$				
	$P_{ik}$ : VH	H	M	L	VL
VH	VH	VH	H	H	M
H	VH	H	H	M	L
M	H	H	M	L	VL
L	H	M	L	VL	VL
VL	M	L	VL	VL	VL

(b) Fuzzy rules of preemption impact intensity calculation					
$x_j$	$I_{pi}$				
	$v_{dj}$ : VH	H	M	L	VL
VH	VH	VH	H	M	M
H	VH	H	M	L	L
M	H	M	M	L	VL
L	M	L	L	VL	VL
VL	L	VL	VL	VL	VL

(c) Fuzzy rules of emergency vehicle preemption priority classification					
$I_{pd}$	$D_p$				
	$I_{pi}$ : VH	H	M	L	VL
VH	VH	VH	VH	VH	VH
H	H	H	VH	VH	H
M	M	H	VH	H	M
L	L	L	M	L	VL
VL	VL	VL	L	L	VL

so the distance is negligible. Total quantity of vehicles on road section  $k$  can be calculated as follows:

$$n_k = \frac{V_k L_k}{N_k v_k}. \quad (3)$$

In (3),  $n_k$  is total quantity of vehicles on road section  $k$ . Parameter  $V_k$  is the current volume of section  $k$ . Parameter  $L_k$  is length of section  $k$ . Parameter  $N_k$  is the number of lanes in section  $k$ . Parameter  $v_k$  is mean travel speed of normal vehicles in section  $k$ . And quantity of vehicles that leave the lane in unit time  $n'_k$  is calculated as follows:

$$n'_k = \beta \mu L_k \frac{N_k}{V_k}. \quad (4)$$

In (4), Parameter  $\beta$  is lane clear influence coefficient that represents the influence of different lane-clear methods on lane clear speed. Its value is defined as 1 when the lane-clear method is changing the lane from inside lane to outside lane in the two-lane sections. The value is demarcated by following simulation experience. And parameter  $\mu$  is the ratio of lane-clear speed to mean time-headway. The value is demarcated by following simulation experience.

Then we can find quantity of vehicles on the emergency lane after normal vehicle avoiding emergency vehicle  $n_{ek}$  is calculated as follows:

$$n_{ek} = \max(n_k - t_k n'_k, 0) = \max\left(\frac{V_k L_k}{N_k v_k} - t_k \beta \mu L_k \frac{N_k}{V_k}, 0\right). \quad (5)$$

In (5),  $t_k$  is road section clearance time. It is a function of emergency vehicle degree of priority  $D_{pk}$ . The formula is shown in (4). This formula should be calibrated based on field data in application. Consider the following:

$$t_k = f(D_{pk}). \quad (6)$$

The applications and studies of road section travel function are the most popular methods to estimate the emergency vehicle travel time by free flow speed, traffic volume, or capacity of the section. This study proposed an emergency vehicle travel time estimation model for the emergency vehicle control method based on bureau of public roads function (BPR function) [31]. BPR function was proposed by FHWA in 1964 in the US [32]. It is the most widely used impedance function in the traffic-planning field. In this function, travel time is the nonlinear function using the ratio of traffic volume to capacity as the parameter.

The definition of the public roads function is defined as follows:

$$T_k = T_0 \left[ 1 + a \left( \frac{V_k(t)}{C_k(t)} \right)^b \right]. \quad (7)$$

In (7),  $T_k(t)$  is emergency vehicle estimated travel time in road section  $k$  at  $t$ . Parameter  $T_0$  is the travel time when the volume of section  $k$  is zero. Parameter  $V_k(t)$  is the volume of section  $k$  at  $t$ . Parameter  $C_k(t)$  is the capacity of section  $k$  at  $t$ . Parameters  $a$  and  $b$  are coefficients whose advised values are  $a = 0.15$  and  $b = 4$ .

Based on above BPR function, this study built the emergency vehicle travel time estimation model for the sections that are applied to different kinds of control methods as follows:

$$T_k = \frac{L_k}{v_e} \left[ 1 + a \left( \frac{n_{ek} N_k V_k}{L_k C_k} \right)^b \right]. \quad (8)$$

In (8),  $n_{ek}$  can be calculated by (5), then, one can get the following:

$$T_k = \frac{L_k}{v_e} \left[ 1 + a \left( \frac{\max(V_k^2 L_k - \Delta t_k \beta \mu N_k^2 v_k, 0)}{V_k L_k C_k} \right)^b \right]. \quad (9)$$

In the formulation, (8) and (9),  $T_k$  is emergency vehicle estimated travel time in section  $k$ . Parameter  $V_k$  is current volume of section  $k$ . Parameter  $L_k$  is length of section  $k$ . Parameter  $N_k$  is the number of lanes in section  $k$ . Parameter  $v_k$  is mean travel speed of normal vehicles in section  $k$ . Parameter  $v_e$  is mean travel speed of emergency vehicles. Parameter  $\beta$  is lane clear influence coefficient.

Parameter  $\mu$  is the ratio of lane-clear speed to mean time-headway. Parameter  $\Delta t_k$  is road section clearance time of section  $k$ . Parameters  $a$  and  $b$  are coefficients demarcated by following simulation experience.

If road section clearance time  $\Delta t_k$  takes zero, that is to say, the emergency vehicle control method is disabled,  $T_k$  can be calculated by the following equation:

$$T_k = \frac{L_k}{v_e} \left[ 1 + a \left( \frac{V_k}{C_k} \right)^b \right]. \quad (10)$$

## 5. Development of Emergency Vehicle Preemption Strategies

**5.1. Route Selection Parameter Calculation and Optimal Route Selection.** In order to select the optimal route for emergency vehicle, route selection parameter should be calculated. The route selection method should provide an efficient and safe traveling route for emergency vehicle with minimum disruption on network traffic. And the disruption on network traffic of emergency vehicle declines with declining degree of priority. Therefore, the model has two optimization objectives including reducing route travel time to minimum and controlling the sum of degree of priority in the route to minimum. In this study, route selection parameter  $P_k$  is defined as follows:

$$P_k = T_k + \eta D_{pk}. \quad (11)$$

In (9),  $T_k$  is emergency vehicle estimated travel time in road section  $k$ .  $D_{pk}$  is emergency vehicle degree of priority in road section  $k$ . Parameter  $\eta$  is degree of priority adjusting coefficients which will be determined in specific case.

Once the location of an emergency issue and emergency facilities (police stations, aid stations, etc.) is determined, the route-selection algorithm determines the best route that has the minimum route selection parameter for a given origin/destination pair. In the proposed strategy, a network is represented as a set of links/nodes and the well-known shortest-path algorithm developed by Dijkstra [30] is adopted to find the optimal route. Dijkstra's algorithm has been proven to result in the shortest-path from a single source on a weighted directed graph, where all edge weights have non-negative values [33]. In the proposed strategy, a given network is modeled as a set of directional links with nonnegative route selection parameter  $P_k$ , and Dijkstra's algorithm is applied to find the minimum  $P_k$  route.

**5.2. Control Method Selection and Optimization.** As soon as the emergency route is determined, this study proposes a strategy to select the control method based on degree of priority. Experience-based control method choice suggestion is shown in Table 2.

When degree of priority of the emergency vehicle is within 0.0 to 0.4, no control method will be implemented to minimize negative effects of emergency vehicle on normal network traffic flow. And if degree of priority is within 0.4 to 0.7, green-wave signal control method will be implemented to reduce travel time of emergency vehicle. Further, when

TABLE 2: Emergency vehicle control method choice suggestion.

Degree of priority	Control method
0.0-0.4	None
0.4-0.7	Green-wave signal preemption
0.7-0.9	Lane clearance preemption
0.9-1.0	Road clearance preemption

degree of priority is within 0.7 to 0.9, the degree of priority is high enough to implement lane clearance control method additionally based on green-wave signal control method. At last, if degree of priority is within 0.9 to 1.0, related road sections will be temporarily closed to reduce travel time of emergency vehicle.

Determining the right offset times to activate the green signal for the intersections along the emergency route is of critical importance in reducing travel time of emergency vehicles and minimizing negative effects of signal preemption on normal traffic flow. As soon as the road sections that implement green-wave signal control method are determined for a given emergency vehicle, the signal offset time of related intersections is adjusted to ensure unidirectional green wave preemption. We record the target offset between downstream intersection  $C_{dk}$  and upstream intersection  $C_{uk}$  of road section  $k$  as  $O_k$ . Then offset  $O_k$  can be calculated as follows:

$$O_k = T_k \bmod C_{dk}. \quad (12)$$

Record the original offset between downstream intersection  $C_{dk}$  and upstream intersection  $C_{uk}$  as  $O'_k$ . So the offset adjustment recorded as  $\Delta O_k$  is calculated as follows:

$$\Delta O_k = O_k - O'_k. \quad (13)$$

If  $C_{dk} < \sum_{i=0}^k T_i$ , offset should be adjusted per cycle;  $\Delta O_{adjk}$  is calculated as follows:

$$\Delta O_{adjk} = \frac{\Delta O_k}{\left[ \sum_{i=0}^k T_i / C_{dk} \right]}. \quad (14)$$

Else, offset should be adjusted per cycle  $\Delta O_{adjk} = 0$ .

When the emergency vehicle departs, change the cycle of each downstream intersection to  $C_{dk} - \Delta O_{adjk}$  until the offset is equal to target offset.

Lane clearance control method is an important method to reduce travel time of emergency vehicle further. Emergency lane is enabled by traffic police or variable lane control system when the emergency vehicle reaches last road section. Startup time of emergency lane is determined by road section travel time of emergency vehicle.

## 6. Performance Evaluation

Finally the proposed degree-of-priority based control strategy is evaluated in a part of road network in Xicheng district, Beijing, China. Figure 4 is the diagram of the road network.

This area adopted SCOOT signalized intersections control system. Network traffic parameters and network data are shown in Table 3.

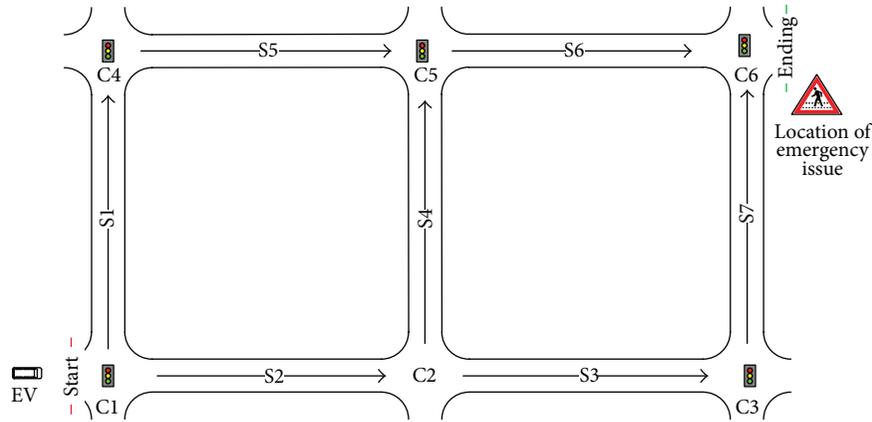


FIGURE 4: The simulation model of road network in Beijing.

TABLE 3: Network parameters and data.

Section name	Section number	Length (m)	Road grade	Traffic volume (PCU/h)		Capacity (PCU/h)
				Direction of arrow	Reversed direction	
Naoshikou Rd.	S1	517	Arterial road	1087	896	3200
West Xuanwumen Rd.	S2	445	Arterial road	684	768	2400
West Xuanwumen Rd.	S3	480	Arterial road	589	656	1600
Tonglinge Rd.	S4	485	Branch road	102	140	600
Xinwenhua Rd.	S5	442	Secondary trunk road	348	268	800
Xinwenhua Rd.	S6	477	Secondary trunk road	248	375	800
Xuanwumen Rd.	S7	463	Arterial road	1395	1674	3200

In this case, fire occurs in a restaurant near intersection C6. Emergency vehicle that is composed of three fire trucks starts from fire station near intersection C1. To reduce loss, the desired travel time of emergency vehicle is set at 180 s. The estimated loss is 2,000,000 CNY, and two of the injured are waiting to be rescued. A microsimulation model based on VISSIM is set up to calibrate some parameters in the strategies. The main processes of simulation are shown in Figure 5.

At a random time, the section in the simulator entered into emergency state which is shown in Figure 5(1). When the section was under emergency state, the vehicles on emergency lane changed to outside lane to clear a lane for the emergency vehicle. Then the emergency vehicle passed the intersection quickly under intersection signal control strategy. The former process is shown in Figures 5(2), 5(3), 5(4), and 5(5). According to the simulation, parameters of control strategy in this case are shown in Table 4.

Based on the degree-of-priority based control strategy and the data in Table 4, degree of priority and emergency vehicle travel time in each road section are calculated by Beijing Special Control System, which is developed by J+traffic Tech Co. The results are shown in Table 5.

By using shortest-path algorithm developed by Dijkstra, the emergency routes based on different  $\eta$  are shown in Figures 6(b) and 6(c).

In Figures 6(b) and 6(c) the red line indicates the road sections that adopt none of the control method, and the green line indicates the road sections that adopt green-wave signal control method. In this study, these two plans are evaluated in simulation software based on cellular automata developed by authors. The degree-of-priority based control strategy is compared with the local-detection based method in the software based on the same set of vehicle inputs. In the microsimulation traffic network, emergency vehicle detectors are set at 100 m from the intersection stop line. For a fair comparison, a common set of 15 different random seeds was used for the simulation of each preemption route, and their results were averaged. The simulation results are shown in Table 6.

As shown in Table 6, the emergency vehicle travel time under degree-of-priority based control strategy exhibits an indistinctive degraded performance. However, a clear pattern of the degraded performance in terms of average delay is exhibited under the degree-of-priority based control strategy. Along with degree of priority adjusting coefficient  $\eta$  increase, thus the selected route is different, the average delay is declining in evidence. This indicates the efficiency of the proposed strategy by reducing unnecessary preemption in a given network, thus minimizing the delay because of preemption without decreasing much efficiency.

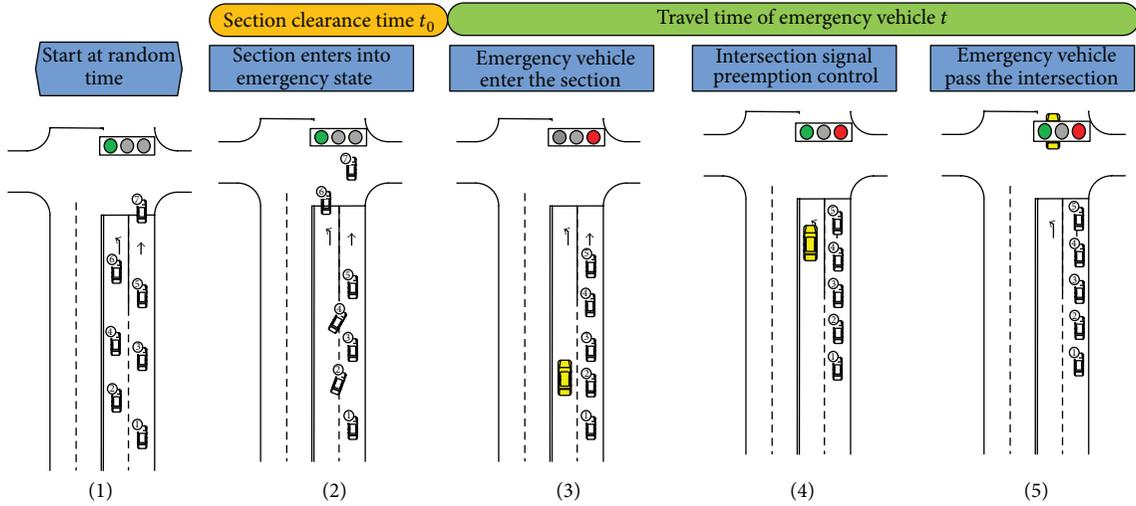


FIGURE 5: Main processes of simulation based on VISSIM.

TABLE 4: Parameters of control strategy in evaluation.

Parameters	Symbol	Value
Conversion coefficient of casualties	$L_k^{(P)}$	70
Road section clearance time	$\Delta t_k$	When $D_p < 0.4$ $\Delta t_k = 3$ When $0.4 < D_p < 0.7$ $\Delta t_k = 6$ Else $\Delta t_k = 20$
Lane clear influence coefficient	$\beta$	1
Ratio of lane-clear speed	$\mu$	135.7
Coefficients of BPR function	$a$	1.96
Coefficients of BPR function	$b$	1.59
Degree-of-priority adjusting coefficients	$\eta$	[10, 50, 100, 200]

TABLE 5: Calculation results of degree-of-priority and emergency vehicle travel time.

Section number	$x_j$	$I_{pi}$	$I_{pd}$	$D_p$	$\Delta t_k$	$T_k$	$\eta$			
							1	50	100	200
S1	0.34	0.48	0.17	0.41	6.0	31.0	35.1	51.6	72.2	113.4
S2	0.29	0.39	0.17	0.39	3.0	26.7	30.6	46.1	65.4	104.1
S3	0.37	0.54	0.17	0.41	6.0	28.8	32.9	49.2	69.5	110.2
S4	0.17	0.21	0.17	0.25	3.0	58.2	60.7	70.5	82.8	107.4
S5	0.44	0.37	0.17	0.37	3.0	40.2	44.0	58.8	77.3	114.4
S6	0.31	0.25	0.17	0.25	3.0	43.1	45.6	55.6	68.1	93.1
S7	0.44	0.62	0.17	0.36	3.0	28.0	31.7	46.2	64.3	100.6

TABLE 6: Simulation results from different preemption strategies.

Routes	Item	Local-detection based method	Degree-of-priority based preemption
Route 1	EV travel time (sec)	103	106
	Total vehicle hours	11.3	10.8
	Average delay (sec)	19.6	15.2
Route 2	EV travel time (sec)	133	142
	Total vehicle hours	10.9	10.6
	Average delay (sec)	16.4	12.7
Route 3	EV travel time (sec)	159	—
	Total vehicle hours	11.1	—
	Average delay (sec)	19.1	—

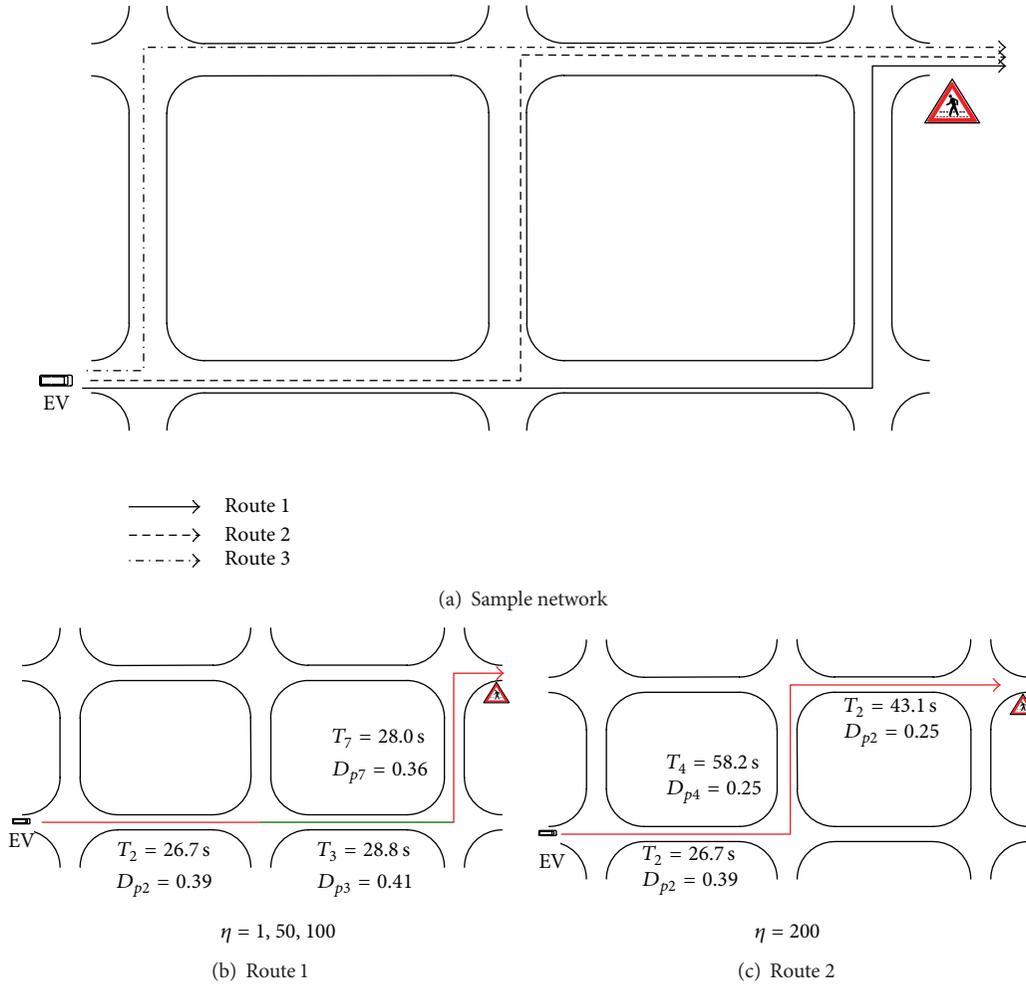


FIGURE 6: Sample network and selected emergency routes based on different  $\eta$ .

## 7. Conclusions

The primary objective of this paper is developing a degree-of-priority based control strategy for emergency vehicle operation to decrease the adverse impacts of emergency vehicles on normal traffic as well as avoid the grid-lock caused by emergency issues. A uniform optimization structure is designed which is consisted by a fuzzy logic based model that was proposed to determine the degree-of-priority of an emergency event. Then, an optimal emergency route selection model was built based on degree-of-priority and the estimated travel time of emergency vehicle. Four types of signal control methods were proposed to minimize the delay of emergency vehicles as well as adverse impacts on normal traffic. Therefore, the traffic signals at the intersections along the emergency route can be adjusted effectively in advance, and traffic queue at each intersection can be cleared for the approaching emergency vehicles. This strategy is evaluated in simulation software based on the data of a part of road network in Xicheng district, Beijing, China. The results indicate that the proposed approach is a viable addition to local-detection based method in urban area and can

minimize the delay of emergency vehicles without much adverse effects on normal traffic. Future research includes the development of an efficient calibration method for the degree of priority with parameters which can be measured and the enhancement of the simulation model to evaluate the performance of the strategy in large road network.

## Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was made possible through funding from the Ministry of Science and Technology of China under Grant no. 2011AA110305. The authors wish to acknowledge the vast and helpful assistance from Intelligent Transportation System Research Center Transportation System Engineering and ITS Research Group, School of Transportation Engineering, Tongji University. The Beijing Special Control System was established with the great help of J+ traffic Tech Co., Ltd.

## References

- [1] Transportation Research Board (TRB), *Proceedings of National Conference on Traffic Incident Management*, 2002.
- [2] U. S. Department of Transportation (DOT), *Fatality Analysis Reporting System*, NHTSA, Washington, DC, USA, 2002.
- [3] Federal Highway Administration (FHWA), *Traffic Signal Preemption for Emergency Vehicles: A Cross-Cutting Study*, 2006.
- [4] D. Deeter, H. M. Zarean, and D. Register, "Rural ITS Toolbox, Section 3: Emergency Services, Subsection 3. 1 Emergency Vehicle Traffic Signal Preemption," 2001.
- [5] D. Bullock, J. Morales, and B. Sanderson, "Evaluation of emergency vehicle signal preemption on the Route 7 Virginia corridor," FHWA-RD-99-70, Federal Highway Administration, Richmond, Va, USA, 1999.
- [6] D. Bullock and E. Nelson, "Impact evaluation of emergency vehicle preemption on signalized corridor operation," in *Proceedings of the TRB Annual Meeting*, Transportation Research Board, Washington, DC, USA, January 2000.
- [7] K. Hunter-Zaworski and A. Danaher, "NE Multnomah street Opticom bus signal priority pilot study," Final Report TNW97-03, Seattle, Wash, USA, 1995.
- [8] Traffic Technologies LLC, "Sonem 2000 digital siren detector," Technical Document for Installation, Traffic Technologies, LLC, 2001.
- [9] E. Kwon, K. Sangho, and B. Rober, "Route-based dynamic preemption of traffic signals for emergency vehicles operations," in *Transportation Research Record*, Transportation Research Board of the National Academies, Washington, DC, USA, 2003.
- [10] C. Louisell and J. Collura, "A simple algorithm to estimate emergency vehicle travel time savings on preemption equipped corridors: a method based on a field operational test," in *Transportation Research Record*, Transportation Research Board of the National Academies, Washington, DC, USA, 2005.
- [11] R. Mussa and M. F. Selekwa, "Development of an optimal transition procedure for coordinated traffic signals," *Journal of Advances in Transportation Studies*, vol. 3, pp. 57–70, 2004.
- [12] A. Haghani, H. J. Hu, and Q. Tian, "An optimization model for real-time emergency vehicle dispatching and routing," in *Transportation Research Record*, Transportation Research Board of the National Academies, Washington, DC, USA, 2003.
- [13] I. Yun, B. B. Park, C. K. Lee, and Y. T. Oh, "Comparison of emergency vehicle preemption methods using a hardware-in-the-loop simulation," *KSCE Journal of Civil Engineering*, vol. 16, no. 6, pp. 1057–1063, 2012.
- [14] I. Yun, B. B. Park, C. K. Lee, and Y. T. Oh, "Investigation on the exit phase controls for emergency vehicle preemption," *KSCE Journal of Civil Engineering*, vol. 15, no. 8, pp. 1419–1426, 2011.
- [15] Q. He, K. L. Head, and J. Ding, "Heuristic algorithm for priority traffic signal control," *Transportation Research Record*, vol. 2259, pp. 1–7, 2011.
- [16] P. T. Savolainen, T. K. Datta, I. Ghosh, and T. J. Gates, "Effects of dynamically activated emergency vehicle warning sign on driver behavior at urban intersections," *Transportation Research Record*, vol. 2149, pp. 77–83, 2010.
- [17] T. S. Glickman and E. Erkut, "Assessment of hazardous material risks for rail yard safety," *Safety Science*, vol. 45, no. 7, pp. 813–822, 2007.
- [18] M. Davidich and G. Köster, "Towards automatic and robust adjustment of human behavioral parameters in a pedestrian stream model to measured data," *Safety Science*, vol. 50, no. 5, pp. 1253–1260, 2012.
- [19] F. Ozel, "Time pressure and stress as a factor during emergency egress," *Safety Science*, vol. 38, no. 2, pp. 95–107, 2001.
- [20] S. Heliövaara, J.-M. Kuusinen, T. Rinne, T. Korhonen, and H. Ehtamo, "Pedestrian behavior and exit selection in evacuation of a corridor—an experimental study," *Safety Science*, vol. 50, no. 2, pp. 221–227, 2012.
- [21] J.-B. Sheu, "An emergency logistics distribution approach for quick response to urgent relief demand in disasters," *Transportation Research E*, vol. 43, no. 6, pp. 687–709, 2007.
- [22] Y. Deng and F. T. S. Chan, "A new fuzzy Dempster MCDM method and its application in supplier selection," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9854–9861, 2011.
- [23] Y. Deng, R. Sadiq, W. Jiang, and S. Tesfamariam, "Risk analysis in a linguistic environment: a fuzzy evidential reasoning-based approach," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15438–15446, 2011.
- [24] A. Hatami-Marbini, M. Tavarna, M. Moradi, and F. Kangi, "A fuzzy group electre method for safety and health assessment in hazardous waste recycling facilities," *Safety Science*, vol. 51, no. 1, pp. 414–426, 2013.
- [25] L. Özdamar and C. S. Pedamallu, "A comparison of two mathematical models for earthquake relief logistics," *International Journal of Logistics Systems and Management*, vol. 10, no. 3, pp. 361–373, 2011.
- [26] H. Rakha and Y. Zhang, "Sensitivity analysis of transit signal priority impacts on operation of a signalized intersection," *Journal of Transportation Engineering*, vol. 130, no. 6, pp. 796–804, 2004.
- [27] F. Dion, H. Rakha, and Y. Zhang, "Evaluation of potential transit signal priority benefits along a fixed-time signalized arterial," *Journal of Transportation Engineering*, vol. 130, no. 3, pp. 294–303, 2004.
- [28] R. J. Baker, J. Collura, J. J. Dale et al., *An Overview of Transit Signal Priority-Revised and Updated*, ITS America, Washington, DC, USA, 2004.
- [29] X. Qin and A. M. Khan, "Control strategies of traffic signal timing transition for emergency vehicle preemption," *Transportation Research C*, vol. 25, pp. 1–17, 2012.
- [30] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.
- [31] H. Spiess, "Conical volume-delay functions," *Transportation Science*, vol. 24, no. 2, pp. 153–158, 1990.
- [32] U.S. Department of Transportation Federal Highway Administration, *Highway History*, 2011, <http://www.fhwa.dot.gov/highwayhistory/nepa/01.cfm>.
- [33] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, Mass, USA, 2nd edition, 2001.

## Research Article

# Optimal Design of the Feeder-Bus Network Based on the Transfer System

Lianbo Deng, Wei Gao, Yanbing Fu, and Wenliang Zhou

*School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China*

Correspondence should be addressed to Yanbing Fu; [quanshuiqq@163.com](mailto:quanshuiqq@163.com)

Received 25 August 2013; Accepted 25 October 2013

Academic Editor: Huimin Niu

Copyright © 2013 Lianbo Deng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper studied the classic feeder-bus network design problem (FBNDP), which can be described as follows: for the passenger travel demand between rail stations and bus stops on a given urban transit network, it designs the optimal feeder bus routes and frequencies so as to minimize the passengers' travel expense and the operator's cost. We extended the demand pattern of M-to-1 in most existing researches to M-to-M. We comprehensively considered the passenger travel cost, which includes the waiting and riding cost on the bus, riding cost on rail, and transfer cost between these two transportation modes, and presented a new genetic algorithm that determines the optimal feeder-bus operating frequencies under strict constraint conditions. The numerical examples under different demand patterns have been experienced and analysed, which showed the robustness and efficiency of the presented algorithm. We also found that the distribution pattern of the travel demand has a significant influence on the feeder-bus network construction.

## 1. Introduction

As the two main transport modes in an urban transit system, the rail line usually plays the role of the transport trunk, while the feeder-bus network services act as a branch of and a supplement to the former. The integration and coordination of urban rail transit and the bus network can effectively promote the service efficiency and simultaneously improve the financial status of the system [1]. Stanger and Vuchic [2] pointed out that coordinative schedule optimization of the two modes could lead to operating cost savings. Some cities, such as Atlanta, Miami, and Washington, DC, gave top priority to the bus/rail coordination during the development process of the transportation systems. Dunn Jr. [3] showed that the coordination and integration of transport services have been a precondition for improving public transportation.

A good feeder-bus network significantly improves the public transport system's service level, operation efficiency, and market competitiveness. The feeder-bus network transports transfer passengers who come from the urban perimeter zone and will arrive at their final business or work destination by rail. Each bus line in the feeder-bus system

usually connects to a special railway station and serves a sequence of bus stops with a certain frequency. Thus, the feeder-bus network design problem (FBNDP) can be described as follows: for a given urban rail line, the stop locations and the passenger travel demand between bus stops and railway stations, the optimal feeder bus routes, and their frequencies are determined so as to minimize the passenger travel cost and the bus operation cost [4–6].

## 2. Literature Review

The existing research on the FBNDP mainly follows two approaches, that is, the analytic approach and network programming (also known as mathematical programming). Most early research used analytic approaches to deduce the optimal route spacing, operating headway, and stop spacing based on assumptions regarding the shape of the street geometry and the spatial distribution of the passenger demand. According to the assumption of the early research, the demand is distributed in a rectangular region in which an existing rail line is serviced (accessed) by some parallel bus routes perpendicular to the rail line. Byrne and Vuchic [4] studied the optimal location and headway of parallel bus lines

and presented a method to determine the optimal number of bus lines. On the basis of [4], Byrne [7] determined the lengths, positions, and headways of bus lines that could minimize the user travel time and operating costs in response to a general population density function with differing line speeds. Hurdle [8] studied the optimal location and schedule of parallel feeder lines with variable passenger density with different trip origins and times. Wirasinghe et al. [9] put forward optimization formulations for the optimal railway interstation spacing, feeder-bus zone boundary, and train headways mainly by the use of basic calculus in conjunction with continuum approximations of certain discrete parameters. Wirasinghe [10] researched a feeder-bus system with a demand pattern of M-to-1 (i.e., multiple bus stops and a single station). An approximate analytical model and corresponding solution algorithm were successfully applied to the Calgary (South Corridor) LRT system. Kuah and Perl [5] optimized the route spacing, operating headway, and stop spacing simultaneously, and analysed the influencing factors of bus stop spacing in three different cases. Supposing that the location of the rail line was predetermined, Chien and Schonfeld [11] cut the urban corridor into several traffic zones with different lengths but the same width and jointly optimized the rail line length, railway station spacing, bus headways, bus stop spacing, and bus route spacing under the conditions that the passenger flow density in each traffic zone was the same and that only one feeder-bus line connected to the same railway station. Chien and Yang [12] developed a model for finding the optimal bus route location and its operating headway in a heterogeneous service area while considering intersection delays. In these models, irregular and discrete M-to-1 demand distributions were considered. A heuristic algorithm [12] and later a genetic algorithm [13] were designed to solve the above model.

In recent decades, the network programming approach has been introduced to deal with the FBNDP. In this approach, the urban transport network is usually represented by a graphic framework, in which nodes denote bus stops or railway stations and links denote route segments between the two successive nodes. For simplicity, it takes bus stops as the origin and the rail station as the destination of the travel demand. Kuah and Perl [6] developed a mathematical programming model for the FBNDP under the M-to-1 demand pattern and designed a heuristic algorithm based on the savings approach. The demand pattern of M-to-M (i.e., the demand pattern in which multiple stations are the destinations) was transformed and generalized to M-to-1 by separating the bus stops into dummy child nodes with the same number of railway stations. Furthermore, the sensitivity of the model was analysed for changes in the design objective, passenger demand variability, vehicle capacity, labour and fuel costs, and rail line. Martins and Pato [14] further presented two strategies to generate the initial solution (i.e., the continuous construction method and the two-phase method) and designed a local search as well as tabu search heuristics with diversification and intensification strategies. Shrivastav and Dhingra [15] discussed the FBNDP for the operational integration of the suburban railway and bus transit system and developed a heuristic algorithm using different node

selection and insertion strategies. Kuan et al. mainly focused on the application of metaheuristic algorithms to the FBNDP, such as simulated annealing and tabu search [16], genetic algorithm, and ant colony optimization [17] and analysed and compared the optimal results obtained by these algorithms.

More recently, Ciaffi et al. [18] dealt with the FBNDP using a two-phase method. In the first phase, a heuristic algorithm was used to generate two different and complementary sets of feasible routes, in order to provide a proper balance between the maximization of the service coverage area and the minimization of the overall travel time. In the second phase, the sets generated in the first phase were used as input data and a GA was designed to find a suboptimal set of routes with the associated frequencies.

Almost all the existing research assumed that the travel demand starts from multiple bus stops but ends at one particular railway station near the central business district (i.e., the demand pattern of M-to-1). In this paper, this drawback is modified to adapt well to the realistic passenger distribution. We consider that origin-destination (OD) pairs may exist between any bus stop and any railway station (M-to-M). After computing the passenger travel cost from the view of the whole feeder-bus network, we present a network optimization model with the objective of minimizing the passenger travel cost and the bus operation cost. Furthermore, a new generation algorithm (GA) is developed and the optimal results under different passenger patterns are analysed and compared.

### 3. Problem Description

The feeder-bus network mainly transports transfer passengers between the bus and the railway system. We regard the nodes, including bus stops and rail stations, as traffic points of passenger collection and distribution. In the classical FBNDP, all passengers are supposed to have a certain railway station as their destination. We widen this assumption to the M-to-M pattern; that is, the passenger demand is distributed between any bus stop and any railway station. Under this demand pattern, the feeder-bus routes obey the following assumptions in most previous studies.

- (1) Each bus stop is served by one feeder-bus route only.
- (2) Each bus route does not cross its feeder railway station but terminates at the station.
- (3) All bus routes have a uniform capacity and operating speed, and the volume of transport passengers should not exceed their capacity.
- (4) Each bus must halt at all the stops along its route and the skip-stop running strategy is not considered.

When the feeder-bus network obeys the above assumptions, there is an M-to-1 connection relationship between bus stops and railway stations. However, the route structure and the feeder station are influenced by the demand distribution of the M-to-M demand pattern, and the operating frequency can be obtained through the optimal cost of the transit operator and transfer passengers.

Considering an urban public transit network composed of  $I$  bus stops and  $J$  railway stations, we denote the set of bus stops by  $B = \{1, \dots, I\}$ , the set of railway stations by  $T = \{I + 1, \dots, I + J\}$ , and the set of network nodes by  $N = B \cup T$ . We also let the distance of a section between two adjacent nodes  $i, j$  be  $L_{ij}$ ,  $i, j \in N$ , the bus operating speed  $v_B$ , and the train speed  $v_T$ .

Due to the time fluctuation of the urban passenger demand, we can determine the bus schedule for each planning period (such as one hour in the morning peak). In the given period, the demand between  $i$  and  $d$  can be expressed as  $P_{id}$ , for  $i \in B, d \in T$ .

Supposing the feeder-bus network  $\Omega$  consists of  $K$  feeder-bus routes, the path structure of bus route  $k$  ( $k = 1, \dots, K$ ) is denoted by  $\omega_k = \{n_1^k, n_2^k, \dots, n_{p_k-1}^k, n_{p_k}^k\}$ , in which  $n_{p_k}^k \in T$  is the feeder railway station and  $n_1^k, n_2^k, \dots, n_{p_k-1}^k \in B$  are bus stops along the route. The operating frequency of route  $k$  is denoted by  $f_k$ .

## 4. Model Construction

The constraints of the feeder-bus network can be obtained according to the above assumptions and the transit operating requirement. Compared with the demand pattern of M-to-1, the network construction constraints under the M-to-M pattern are completely uniform. However, the generalized travel cost of transfer passengers under the demand pattern of M-to-M will have a more complex influence on their choice of feeder station in the railway line, thus affecting the feeder-bus route structure and further the network construction.

**4.1. Constraints Analysis.** To represent the feeder-bus network constraints,  $Y_{ij}$  and  $X_{ihk}$  are defined to denote the relationship between nodes and routes:

$$Y_{ij} = \begin{cases} 1, & \text{if bus node } i \text{ is assigned to rail node } j \\ 0, & \text{otherwise,} \end{cases}$$

$$\forall i = 1, \dots, I; j = I + 1, \dots, I + J,$$

$$X_{ihk} = \begin{cases} 1, & \text{if node } i \text{ precedes node } h \text{ on bus route } k \\ 0, & \text{otherwise,} \end{cases}$$

$$\forall i, h = 1, \dots, I + J; k = 1, \dots, K. \quad (1)$$

All the constraints that need to be satisfied by the feeder-bus network are as follows.

- (1) Connectedness constraint of the feeder-bus network: in the feeder-bus network, any subset of bus stops must link to feeder stations directly or via other bus stops, that is, the following connectedness constraint:

$$\sum_{i \notin H} \sum_{h \in H} \sum_{k=1}^K X_{ihk} \geq 1, \quad \forall H, \quad (2)$$

where  $H$  is a set containing all the rail stations and some bus stops. Obviously, it is also a subset of  $N$ .

- (2) Integrity constraints of the feeder-bus route:

each bus route must link to a single railway station:

$$\sum_{i=1}^I \sum_{j=I+1}^{I+J} X_{ijk} = 1, \quad k = 1, \dots, K. \quad (3)$$

A route terminates at a certain feeder station  $d \in T$  to which the route passengers are transported,

$$\sum_{i=I+1}^{I+J} \sum_{j=1}^{I+J} X_{ijk} = 0, \quad k = 1, \dots, K. \quad (4)$$

In addition, a feeder-bus route should include at least one stop and one feeder station, that is, the following nonempty constraint:

$$\sum_{i=1}^I \sum_{j=1}^{I+J} X_{ijk} \geq 1, \quad k = 1, \dots, K. \quad (5)$$

- (3) Relationship constraints between routes and nodes:

feeder-bus route  $k$  must halt at stop  $i$  only once,  $i$  should be served only by route  $k$  when  $i \in \omega_k$ , and route  $k$  should be acyclic:

$$\sum_{k=1}^K \sum_{h=1}^{I+J} X_{ihk} = 1, \quad i = 1, \dots, I,$$

$$\sum_{k=1}^K \sum_{p=1}^I X_{pik} \leq 1, \quad i = 1, \dots, I,$$

$$\sum_{h=1}^{I+J} X_{ihk} - \sum_{p=1}^I X_{pik} \geq 0, \quad i = 1, \dots, I; k = 1, \dots, K. \quad (6)$$

In addition, relationship variable  $Y_{ij}$  between route  $k$  and nodes  $i, j$  satisfies the following constraint:

$$\sum_{h=1}^{I+J} X_{ihk} + \sum_{p=1}^I X_{pjk} - Y_{ij} \leq 1, \quad i = 1, \dots, I; \\ j = I + 1, \dots, I + J; k = 1, \dots, K. \quad (7)$$

- (4) Capacity constraints of the feeder-bus route and network:

for route  $k$ , operating frequency  $f_k$  should meet the passenger transport capacity; that is,

$$\sum_{i=1}^I \sum_{d=I+1}^{I+J} P_{id} \sum_{h=1}^{I+J} X_{ihk} \leq f_k C \rho, \quad k = 1, \dots, K, \quad (8)$$

where  $C$  is the bus capacity and  $\rho$  is the load factor.

Meanwhile, the route operating frequencies are restricted by the total running mileage of bus vehicles in the planning period; that is,

$$\sum_{k=1}^K f_k \sum_{i=1}^I \sum_{h=1}^{I+J} L_{ih} X_{ihk} \leq \frac{1}{2} N v_B (T - t), \quad (9)$$

where  $t$  is the average servicing and turn-around time of every bus vehicle in period  $T$  and  $N$  is the number of vehicles that can be used in  $T$ .

**4.2. Cost Analysis.** In order to ensure the good market competitiveness and operation effect, the feeder-bus network needs to consider the benefits both of passengers and of operators. For passengers, this refers to their generalized travel cost, which can be divided into three parts: the waiting and riding cost on the bus, the riding cost on the rail line, and the transfer cost between these two modes. Compared with [6, 14], the transfer cost is taken into account in this paper and the contents of the passenger travel cost are closer to a realistic situation.

In the passenger travel cost, the bus waiting and bus riding costs are  $\lambda_w \sum_{k=1}^K (1/2 f_k) \sum_{d=I+1}^{I+J} \sum_{i=1}^I \sum_{h=1}^{I+J} P_{id} X_{ihk}$  and  $\sum_{k=1}^K \{L_k \sum_{d=I+1}^{I+J} P_{n_1^k d} + (L_k - L_{n_1^k n_2^k}) \sum_{d=I+1}^{I+J} P_{n_2^k d} + \dots + L_{n_{p_k-1}^k n_{p_k}^k} \sum_{d=I+1}^{I+J} P_{n_{p_k-1}^k d}\} \lambda_r / v_B$ , respectively, where  $\lambda_w$  and  $\lambda_r$  are the corresponding monetary cost conversion coefficients,  $P_k$  and  $L_k$  are the total passenger demand and length of route  $k$ ; the transfer cost at the feeder station is closely related to the transfer facilities and train operating density, so the transfer cost for every passenger at station  $j$  ( $j = I + 1, \dots, I + J$ ) can be expressed as  $\lambda_j$ , and the total rail riding cost is  $(\lambda_T / v_T) (\sum_{d=I+1}^{I+J} \sum_{j=I+1}^{I+J} \sum_{i=1}^I P_{id} Y_{ij} L_{jd})$ , where  $\lambda_T$  is the corresponding monetary cost conversion coefficient.

For the bus operator, the operating cost of feeder-bus routes mainly relates to operating mileages and frequencies, which are denoted as  $2\lambda_b \sum_{k=1}^K f_k \sum_{h=1}^I \sum_{i=1}^{I+J} L_{ih} X_{ihk}$ , where  $\lambda_b$  is the cost per carriage per mile.

**4.3. Optimization Model.** Considering the benefits of both the passengers and the operator, the objective function of feeder-bus network programming is to minimize the passengers' generalized travel cost and the operator's cost:

$$\begin{aligned} \min Z & \\ &= \frac{\lambda_T}{v_T} \left( \sum_{d=I+1}^{I+J} \sum_{j=I+1}^{I+J} \sum_{i=1}^I P_{id} Y_{ij} L_{jd} \right) \\ &+ \lambda_w \sum_{k=1}^K \frac{1}{2 f_k} \sum_{d=I+1}^{I+J} \sum_{i=1}^I \sum_{h=1}^{I+J} P_{id} X_{ihk} \\ &+ \left( \sum_{k=1}^K \left\{ L_k \sum_{d=I+1}^{I+J} P_{n_1^k d} + (L_k - L_{n_1^k n_2^k}) \sum_{d=I+1}^{I+J} P_{n_2^k d} \right. \right. \end{aligned}$$

$$\begin{aligned} &\left. + \dots + L_{n_{p_k-1}^k n_{p_k}^k} \sum_{d=I+1}^{I+J} P_{n_{p_k-1}^k d} \right\} \lambda_r \right) \times (v_B)^{-1} \\ &+ \sum_{j=I+1}^{I+J} \lambda_j \left( \sum_{d=I+1}^{I+J} \sum_{i=1}^I P_{id} Y_{ij} - P_{ij} \right) \\ &+ 2\lambda_b \sum_{k=1}^K f_k \sum_{h=1}^I \sum_{i=1}^{I+J} L_{ih} X_{ihk}. \end{aligned} \quad (10)$$

Objective function (10) and constraints (2)~(9) constitute the optimization model of the FBNDP. The optimization model has the following main characteristics.

- (1) The transfer cost is included in the passenger travel cost, so the travel cost on the transfer network can be calculated overall.
- (2) The optimal objective is to minimize the passenger travel cost and transit operating cost, so benefits to users and operators in the public transit system are both considered.
- (3) The demands are not limited to a single destination (M-to-1) and are extended to the distribution between every origin and destination (M-to-M), so the model accords well with a realistic demand distribution.

## 5. Model Solution

The optimization model of the feeder-bus network, with many 0-1 variables and many constraints, is NP hard [14]. It is essentially a route optimization problem in the field of operations research, which is appropriate for solving by some intelligent heuristic algorithms. Thus, in this paper, we present a genetic algorithm for the model. In the generation process of initial solution individuals and new population individuals, we use the following strategy for constructing routes: it firstly generates the connection relations of stops and stations, then finally optimizes the routing path structure and determines the operating frequency. In the process of generating feeder-bus routes, feeder relations and the path of each bus route are determined according to the cost of candidate routes.

**5.1. Optimal Frequency for Each Route.** In feeder-bus network  $\Omega$ , the path structure and cost of a feeder-bus route are not correlated with the other routes; therefore, the operating frequency of each route can be set independently. For feeder-bus route  $\omega_k$ , according to the objective function (10), its cost is

$$\begin{aligned} C_k & \\ &= \frac{\lambda_T}{v_T} \left( \sum_{i=n_1^k}^{n_{p_k-1}^k} \sum_{d=I+1}^{I+J} P_{id} L_{n_{p_k}^k d} \right) + \frac{\lambda_w P_k}{2 f_k} \end{aligned}$$

$$\begin{aligned}
& + \lambda_{n_{p_k}} \left( P_k - \sum_{i=n_1^k}^{n_{p_k}^k-1} P_{in_{p_k}} \right) + 2\lambda_b f_k L_k \\
& + \left( \lambda_r \left\{ L_k \sum_{d=I+1}^{I+J} P_{n_1^k d} + (L_k - L_{n_1^k n_2^k}) \right. \right. \\
& \quad \left. \left. \times \sum_{d=I+1}^{I+J} P_{n_2^k d} + \cdots + L_{n_{p_k-1}^k n_{p_k}^k} \sum_{d=I+1}^{I+J} P_{n_{p_k-1}^k d} \right\} \right) \times (v_B)^{-1}.
\end{aligned} \tag{11}$$

In order to minimize  $C_k$ , the optimal operating frequency without any constraint can be obtained by the first-order optimality condition of  $C_k$  with respect to  $f_k$  in (11) as follows:

$$f_k^* = \frac{1}{2} \sqrt{\frac{\lambda_w P_k}{\lambda_b L_k}}. \tag{12}$$

Then, the optimal operating frequency of  $\omega_k$  under constraint (8) is

$$\bar{f}_k^* = \max \left\{ \frac{1}{2} \sqrt{\frac{\lambda_w P_k}{\lambda_b L_k}}, \frac{P_k}{(C\rho)} \right\}. \tag{13}$$

Under the optimal frequency  $\bar{f}_k^*$ , the minimum route cost of  $\omega_k$  is

$$\begin{aligned}
& C_k^* \\
& = \frac{\lambda_T}{v_T} \left( \sum_{i=n_1^k}^{n_{p_k}^k-1} \sum_{d=I+1}^{I+J} P_{id} L_{n_{p_k} d} \right) + \frac{\lambda_w P_k}{2\bar{f}_k^*} \\
& + \lambda_{n_{p_k}} \left( P_k - \sum_{i=n_1^k}^{n_{p_k}^k-1} P_{in_{p_k}} \right) + 2\lambda_b \bar{f}_k^* L_k \\
& + \left( \lambda_r \left\{ L_k \sum_{d=I+1}^{I+J} P_{n_1^k d} + (L_k - L_{n_1^k n_2^k}) \right. \right. \\
& \quad \left. \left. \times \sum_{d=I+1}^{I+J} P_{n_2^k d} + \cdots + L_{n_{p_k-1}^k n_{p_k}^k} \sum_{d=I+1}^{I+J} P_{n_{p_k-1}^k d} \right\} \right) \times (v_B)^{-1}.
\end{aligned} \tag{14}$$

Thus, the total operating cost of  $\Omega$  is

$$Z(\Omega) = \sum_{k=1}^K C_k^*. \tag{15}$$

Particularly, the optimal operating frequency  $\bar{f}_{ij}^*$  of direct-link route  $\omega_{ij}$ , which directly links stop  $i \in B$  and station

$j \in T$ , is  $\max\{(1/2)\sqrt{\lambda_w \sum_{d=I+1}^{I+J} P_{id}/\lambda_b L_{ij}}, \sum_{d=I+1}^{I+J} P_{id}/(C\rho)\}$ . Then, accordingly, the total operating cost  $DC_{ij}$  of  $\omega_{ij}$  is

$$\begin{aligned}
DC_{ij} & = \frac{\lambda_T}{v_T} \left( \sum_{d=I+1}^{I+J} P_{id} L_{jd} \right) + \frac{\lambda_w \sum_{d=I+1}^{I+J} P_{id}}{2\bar{f}_{ij}^*} \\
& + \frac{\lambda_r L_{ij} \sum_{d=I+1}^{I+J} P_{id}}{v_B} + \lambda_j \left( \sum_{d=I+1}^{I+J} P_{id} - P_{ij} \right) + 2\lambda_b \bar{f}_{ij}^* L_{ij}.
\end{aligned} \tag{16}$$

**5.2. Optimization Strategy for the Route Structure.** In the process of the GA, all routes  $\omega_k$  in the population individuals need to optimize the path structures when the nodes in the routes are changed or the generation is updated. When the stops served by route  $\omega_k$  and the end (feeder station) of  $\omega_k$  are determined, the path structure optimization of  $\omega_k$  can come down to an open vehicle routing problem with one depot (i.e., the feeder station). Thus, we adopt  $2p_k$  iterations of the 2-opt strategy to optimize the route structure, where  $p_k$  is the number of nodes on the route. Because of the high probability that some elite gene segments of the offspring can be inherited from the parents, the fixed iterations of the route structure optimization for every individual of each generation can improve the population quality.

**5.3. Genetic Coding.** In this paper, we use an intuitional style to code the feeder-bus plan. Thus, every node in the network is expressed by a natural number; we also identify the bus stops or railway stations with different number sets. Then, a feeder-bus route would be a number substring ending at a railway station, and the whole coding scheme of the network would be the sequential connection of these route substrings. Note that the coding length of the network plan depends on the total route number designed in the transit system and it is not fixed.

For example, when  $B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$  and  $T = \{11, 12, 13\}$ , a sample feeder-bus network can be expressed as 1 2 3 **11** 4 6 **12** 5 7 **12** 8 9 10 **13**; substrings 1 2 3 **11**, 4 6 **12**, 5 7 **12**, and 8 9 10 **13** stand for 4 feeder-bus routes, respectively, in which the boldface numbers stand for the feeder stations.

Based on objective function (10) and considering the feasibility of constraint (9) simultaneously, the fitness function of individual  $\Omega$  is constructed as follows:

$$F(\Omega) = \frac{1}{\lambda Z(\Omega)} + \lambda \left[ \frac{1}{2} N v_B (T - t) - \sum_{k=1}^K f_k \sum_{i=1}^I \sum_{h=1}^{I+J} L_{ih} X_{ihk} \right], \tag{17}$$

where  $\lambda$  is a penalty factor.

**5.4. Initial Population.** Each feeder-bus network in the initial population is generated one route by one route. Due to that, all the routes end at railway stations, so a feeder-bus station  $j \in T$  can be selected first; then we choose a bus stop  $i \in B$  with a choice probability, insert this stop into an existing route that terminates at node  $j$ , or link it directly with node  $j$  to

generate a new route. In this way, the generation of a network is finished when all the bus stops have been selected.

In order to improve the individual quality of the initial population, we construct a function to evaluate the connecting relationship between a bus stop and a railway station so that the roulette selection method is utilized to compute the selection probability. Let  $DC_{ij}^* = \max_{i \in B} DC_{ij} + \min_{i \in B} DC_{ij}$ ; then, the evaluation function between stop  $i$  and station  $j$  is

$$F_{ij} = DC_{ij}^* - DC_{ij}. \quad (18)$$

For station  $j$ , the selection probability of stop  $i$  is  $F_{ij} / \sum_{i=1}^I F_{ij}$ .

The construction algorithm of a feeder-bus network of the initial population is as follows.

*Algorithm 1.* (1) Let  $B' = B$ , where  $B'$  stands for the set of stops that could be selected to construct the current feeder-bus route.  $\Omega = \emptyset$  is a feeder-bus network and  $k = 0$  is the number of routes that have been generated.

(2) If  $B' = \phi$ , the algorithm terminates. Otherwise, randomly select a station  $j$  from  $T$  with equal probability.

(3) A random number  $R \in (0, 1)$  is created according to uniform distribution. Take a stop  $i$  from  $B'$ , which satisfies  $\sum_{i=1}^{i-1} F_{ij} / \sum_{i=1}^I F_{ij} \leq R < \sum_{i=1}^i F_{ij} / \sum_{i=1}^I F_{ij}$ .

(4) Let  $\omega_{ij}$  be the direct route from  $i$  to  $j$  and  $M$  the number of routes ending at  $j$  in  $\Omega$ . If  $M = 0$ , set  $\Omega = \Omega \cup \{\omega_{ij}\}$ ,  $B' = B' \setminus i$ ,  $k = k + 1$ ; go to (2). Otherwise, for  $m = 1, \dots, M$ , insert  $i$  into the  $m$ th route that ends at  $j$  in  $\Omega$ , form the corresponding network  $\Omega_m$  and let  $\Omega_{M+1} = \Omega \cup \{\omega_{ij}\}$ ; let  $\Omega' = \arg \min\{Z(\Omega_m) \mid m = 1, \dots, M + 1\}$ ; if  $\Omega' = \Omega_{M+1}$ ,  $k = k + 1$ ;  $\Omega = \Omega'$ ,  $B' = B' \setminus i$ ; go to (2).

Let  $n$  be the size of the initial population, which could be reasonably determined according to the values of  $I$  and  $J$ . The initial population is constructed by calling Algorithm 1  $n$  times.

### 5.5. Genetic Operators

(1) *Selection and Replication Operator.* To strengthen the searching ability of the GA, competition and intrusion mechanisms are introduced to construct the parent population. The former is that the population with  $n$  individuals is replicated to form a new population with  $2n$  individuals firstly; then these  $2n$  individuals are divided into  $n$  pairs arbitrarily and the better individuals are preserved by comparing the fitness of each pair. The latter means that  $\alpha n$  new individuals by Algorithm 1 are introduced to replace the  $\alpha n$  worst ones of the current population, where  $\alpha$  is an intrusion ratio. The value of  $\alpha$  is dynamically controlled in the range of  $[\underline{\alpha}, \bar{\alpha}]$ . Set  $\alpha = \min(2 * \alpha, \bar{\alpha})$  when the best solution has not been improved in  $T_\alpha$  times generations; set  $\alpha = \underline{\alpha}$  when the best solution is improved.

(2) *Crossover Operator.* Here, we take two parents to generate two offspring with crossover probability  $P_c$ . In order to ensure that offspring individuals can inherit the elite gene from the parents, gene segments of routes of which the average cost per passenger is lower are chosen and inserted

into the offspring individuals. In individual  $\Omega_l$ , let the node of locus  $i$  be  $V_l^i$ , let the number of routes  $n_l$ , and let the average cost per passenger of  $\omega_k$  be  $AC_k$ .

Choose the two parents  $\Omega_{l_1}$  and  $\Omega_{l_2}$  and generate offspring  $\Omega_l$  based on  $\Omega_{l_1}$ . Firstly, compare the average travel cost per passenger of the two gene segments that start the node  $V_{l_1}^1$ , the first locus of  $\Omega_{l_1}$ , in  $\Omega_{l_1}$  and  $\Omega_{l_2}$ . The better option is to choose  $\Omega_l$  and delete  $V_{l_1}^1$  from the two parents. Then, make the comparison of the gene segments that start the last node in  $\Omega_l$  in the two parents until  $\Omega_l$  terminates at a railway station and a route in  $\Omega_l$  is generated. Then, delete the routes that cannot satisfy constraint (5) or combine the two shortest routes in each parent. With the above method, gene segments of the parents are selected to join the offspring constantly, and the routes of the offspring are constructed one by one until the whole offspring feeder-bus network is formed. The detailed algorithm is described as follows.

*Algorithm 2.* (1) Take the parents  $\Omega_{l_1}$  and  $\Omega_{l_2}$ , and the generated offspring  $\Omega_l$ . Set  $\Omega_{l_1}' = \Omega_{l_1}$ ,  $\Omega_{l_2}' = \Omega_{l_2}$ , and  $i = 1$ .

(2) If  $n_{l_1}' = 0$ , go to (7); otherwise, select the locus  $i' = 1$  and the corresponding route is  $\omega_{k_1}$  in  $\Omega_{l_1}'$ . Set  $V_l^i = V_{l_1}^{i'}$ .

(3) Find locus  $j'$  from  $\Omega_{l_2}'$ , satisfying  $V_{l_2}^{j'} = V_{l_1}^{i'}$ , and the corresponding route is  $\omega_{k_2}$ ; go to (5).

(4) Find locus  $i'$  from  $\Omega_{l_1}'$ , satisfying  $V_{l_2}^{j'} = V_{l_1}^{i'}$ , and the corresponding route is  $\omega_{k_1}$ ; go to (5).

(5) When  $AC_{k_1} \leq AC_{k_2}$ , set  $V_l^{i+1} = V_{l_1}^{i'+1}$ ,  $\Omega_{l_2}' = \Omega_{l_2}' \setminus V_{l_2}^{j'}$ ,  $\Omega_{l_2}' = \Omega_{l_2}' \setminus V_{l_2}^{j'}$ ,  $i = i + 1$ , and if  $V_l^{i+1}$  is not a railway station, go to (3); otherwise,  $n_l = n_l + 1$ ; go to (6).

When  $AC_{k_1} > AC_{k_2}$ , set  $V_l^{i+1} = V_{l_2}^{j'+1}$ ,  $\Omega_{l_1}' = \Omega_{l_1}' \setminus V_{l_1}^{i'}$ ,  $\Omega_{l_1}' = \Omega_{l_1}' \setminus V_{l_1}^{i'}$ ,  $i = i + 1$ , and if  $V_l^{i+1}$  is not a railway station, go to (4); otherwise,  $n_l = n_l + 1$ ; go to (6).

(6) Delete those routes in  $\Omega_{l_1}$  and  $\Omega_{l_2}$  that do not satisfy constraint (5). If all the routes satisfy constraint (5) in  $\Omega_{l_1}'$  or  $\Omega_{l_2}'$ , combine the two shortest routes in the corresponding parent; go to (2).

(7) Optimize the path structure of  $n_l$  routes in  $\Omega_l$ , respectively.

Based on  $\Omega_{l_2}$ , another offspring can be generated in the same way.

Taking the following parents, the crossover operator that generates offspring 1 based on parent 1 is shown as follows:

Parent 1: 1 2 3 11 4 6 12 5 7 12 8 9 10 13

Parent 2: 4 8 12 5 2 11 1 9 3 12 10 6 7 13.

First, node 1 at the first locus in parent 1 is used as the node of offspring 1 at locus 1. Then, gene segments 1-2 in parent 1 and 1-9 in parent 2 are compared; 1-2 are supposed to join offspring 1. Subsequently, 2-3 and 2-11 are compared. Offspring 1 is as below:

Offspring 1: 1 2 11 ...

(3) *Mutation Operator.* A mutation operator with mutation probability  $P_c$  is used to strengthen the global optimization ability of the GA. According to the types of genes randomly selected from one individual, exchange or insertion mutations are made: if the selected gene is a bus stop, it will be randomly inserted into another locus (insertion mutation); if the selected gene is a station, it will be randomly replaced by another station (exchange mutation). To guarantee the quality of the mutation, the acceptance probability of a mutation solution is

$$P = \begin{cases} 1, & \Delta Z < 0, \\ e^{(-\Delta Z/Z_0)}, & \Delta Z \geq 0, \end{cases} \quad (19)$$

where  $Z_0$  is the fitness of the best individual so far and  $\Delta Z$  is the objective difference of the solution mutation before and after.

(4) *Elite Preservation Strategy and Algorithm Termination Rules.* To preserve the elite individuals of parent populations, the worst 4% of offspring individuals are replaced by the same proportion of the best ones. The termination rules of the GA make use of the maximum generations  $T_{\max}$  or the maximum generations  $T_0$  without improving the best solution so far.

## 6. Numerical Examples

The benchmark problem is taken from [6]. The network includes 55 bus stops and 4 railway stations, serving  $2 * 2.5$  square miles. The demand density of each stop per period (one hour) is 200 passengers. The values of the model parameters are shown in Table 1. The GA is designed based on the C# language. In the GA, we take  $n = 120$ ,  $P_c = 0.8$ ,  $P_m = 0.08$ ,  $T_{\max} = 1200$ ,  $T_0 = 100$ ,  $\underline{\alpha} = 0.05$ ,  $\bar{\alpha} = 0.30$ , and  $T_\alpha = 20$ .

6.1. *M-to-1 Demand Pattern.* Railway station 56 is regarded as the central business district of the service area and the destination of all the passengers, so the demand is a distribution pattern between multiple stops and one station. The best feeder-bus network is shown in Figure 1 and Table 2.

Figure 2 shows a change in the objective function when the number of generations increases in the solving process. It illustrates that the GA presented in this paper performs a fast convergence speed.

6.2. *Comparison of Best Solutions.* To compare the optimal solutions with other approaches, we neglect the transfer cost (i.e.,  $\lambda_j = 0$ ) and make the total cost of this paper accord with other methods, including saving heuristics [6], displacement heuristics, basic TS [14], and TS with intensification [16]. However, the bus riding-time cost is roughly approximated by estimating the total passenger-miles in these studies and there are some differences in constraints (8) and (9) in the models of [6, 16]. Table 3 gives the best solutions of these studies.

The results show that the total cost of the GA saves 8.0%, 1.3%, 1.2%, and 1.4% compared to the other approaches [6, 14, 16], respectively. Because the length of each route is

TABLE 1: Model parameters.

Parameter	Unit	Value
$C$	Seat	50
$\rho$		1.2
$t$	Hour	0.16
$N$	Vehicle	110
$\lambda_w$	\$/passenger-hour	8
$\lambda_r$	\$/passenger-hour	4
$\lambda_T$	\$/passenger-hour	4.5
$\lambda_i$	\$/passenger	0.03
$\lambda_b$	\$/vehicle-mile	3
$V_B$	Mile/hour	20
$V_T$	Mile/hour	30
$\lambda$		0.0001

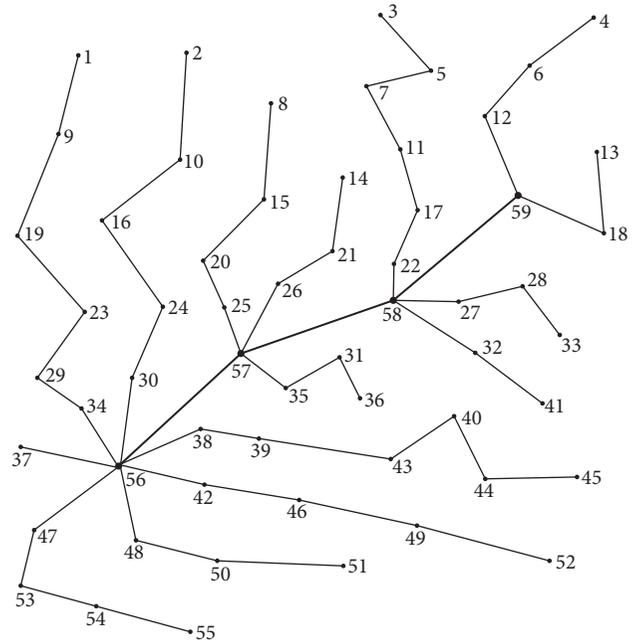


TABLE 3: Comparison of the best solutions.

Solution approaches	Total cost (\$)	Route no.	Average length per route (miles)	Average frequency per route (trips/hour)	Average computational time (seconds)
Saving heuristics [6]	6824	16	1.05	21.06	—
Displacement heuristics [14]	6363	18	0.84	22.54	—
Basic TS [14]	6355	18	0.85	22.18	—
TS with intensification [16]	6371	19	0.80	22.76	13.8
GA (in this paper)	6281	15	1.01	22.01	21.6

TABLE 4: Indicators of optimal solutions under various demand distributions.

$a$	System cost (\$)	Station's transfer passengers				Number of routes	Average route length	Average route frequency	Average travel time (bus : train)	Nonlinear coefficient
		56	57	58	59					
50	6515	1600	3800	4600	1000	19	0.79	21.88	0.51 : 0.53	1.28
40	6511	1400	3600	5000	1000	19	0.80	21.93	0.51 : 0.51	1.27
30	6505	800	3800	4600	1800	18	0.84	22.06	0.50 : 0.48	1.27
20	6494	1000	3000	5200	1800	18	0.84	21.99	0.50 : 0.48	1.25
10	6476	800	2800	5600	1800	16	0.93	22.21	0.47 : 0.41	1.25
0	6437	800	2800	4800	2600	17	0.89	22.04	0.50 : 0.42	1.22

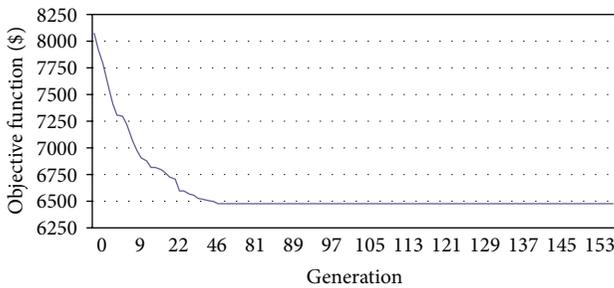


FIGURE 2: Convergence efficiency of GA under M-to-1.

patterns of passenger demand between each stop and 4 railway stations are generated by an arithmetic progression, the first term of which is  $a$  and the common difference is  $q$ . For example, when  $a = 20$  and  $q = 20$ , the number of passengers from each stop to stations 56–59 is 20, 40, 60 and 80, respectively. Obviously, the imbalance of the demand distribution increases when  $a$  decreases from 50 to 0. The results under different values of  $a$  are shown in Table 4 and Figure 3 shows the optimal feeder-bus network under the uniform distribution demand between the 4 railway stations ( $q = 0$ ).

Figure 4 shows a change in the objective function under the demand pattern of M-to-M and  $q = 0$ . It shows that the convergence speed of GA under M-to-M is satisfactory as well as M-to-1. From Table 4, the following observations regarding the demand distributions effect on the optimal feeder-bus networks can be found.

- (1) For a given railway station, with the increase in passengers who terminate at a station, the number of passengers who choose to feed into this station increases grossly, as seen in Figure 5.

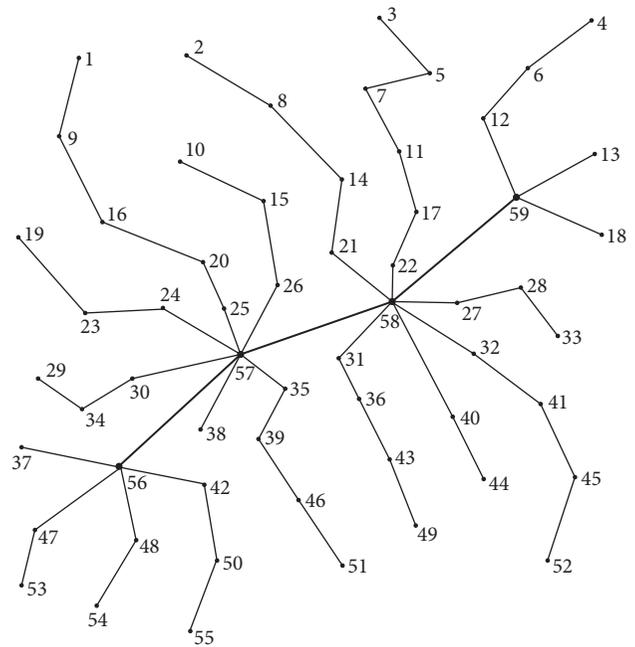


FIGURE 3: Optimal feeder-bus network under  $q = 0$ .

- (2) The demand distribution has an obvious effect on the average riding time by bus and train, as shown in Figure 6. With the increase in the demand imbalance between the stations, the difference in the average travel time in the two traffic modes becomes gradually more significant and the total travel time on the integrated transport network decreases simultaneously because the feeder station and route structure are influenced by major passengers.

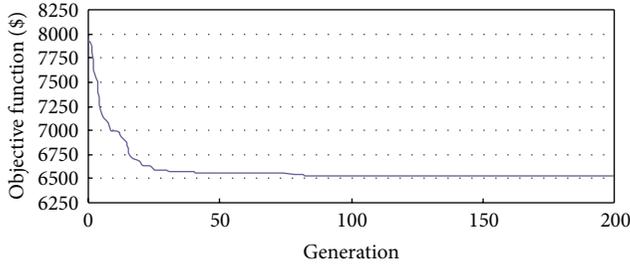


FIGURE 4: Convergence efficiency of GA under M-to-M.

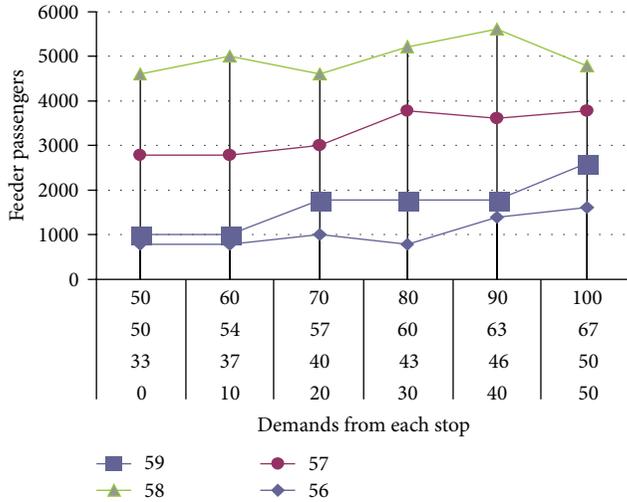


FIGURE 5: Effect of termination passengers on feeder passengers at stations.

(3) As Figure 7 shows, with the increase in the demand imbalance between stations, namely, the concentration of demand destinations, passengers' nonlinear coefficient falls and the system total cost also decreases remarkably. The reason is that the concentration of demand destinations makes most passengers obtain a better service, which causes a decline in the system's total cost.

According to the above analysis and the difference between Figures 1 and 3, we can draw the conclusion that the demand distribution has a great effect on the volume of passengers choosing given feeder-bus stations and the path structure of feeder routes, which will further influence the total cost of the whole feeder system.

### 7. Conclusions

This paper studies the optimal design problem of a feeder-bus network under the demand pattern of M-to-M. The drawback in most existing related research, that only a single destination exists (M-to-1 demand pattern), is modified to the M-to-M pattern for better accordance with the realistic demand distribution. In order to minimize the passenger travel cost and transit operating cost, an integrated public transport system of a feeder-bus network and railway is regarded as a

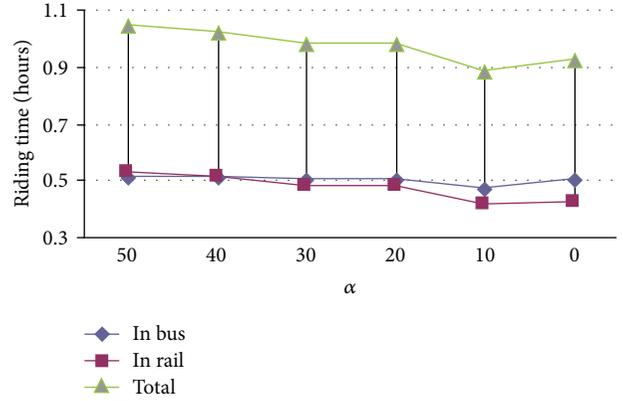


FIGURE 6: Effect of demand distribution on passengers' travel time.

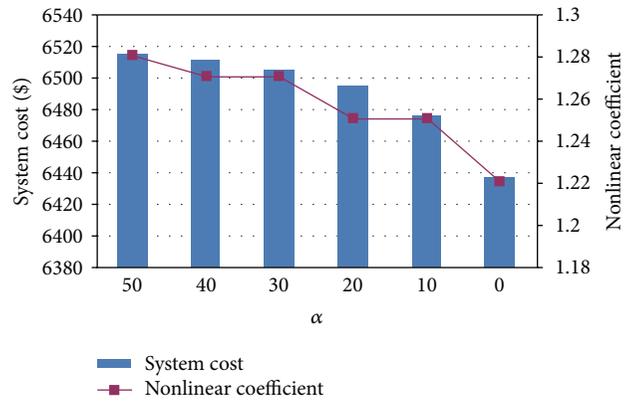


FIGURE 7: Relation between demand distribution and system cost or nonlinear coefficient.

whole to calculate the passenger travel cost overall. The results show that passenger demand distributions have a significant influence on feeder-bus network construction, especially on the feeder stations, the paths and frequencies, of feeder bus routes. Therefore, demand distributions should be considered when designing a feeder-bus network.

Usually public transportation network planning has symmetry, though differences in two directions are not considered in this paper. If the demand on a feeder-bus network has an obvious tidal phenomenon with time distribution, and the operating frequencies in different directions differ greatly, a directed feeder-bus network should be designed according to the directional demand.

One prerequisite of the FBNDP in this paper is that the station layout has to be determined. In our work, the optimal frequencies of feeder bus routes are determined according to the feeder passenger between rail and bus transit system. If one feeder-bus route only services few passengers, it could be removed from the feeder-bus network.

### Acknowledgments

This research is supported by the National Natural Science Foundation of China (70901076, 71171200), the Science and

Technology Research Development Program of China Railway Corporation (Major Program, 2013X004-A), and the Research Fund for Fok Ying Tung Education Foundation of Hong Kong (Project no. 132017).

## References

- [1] A. D. May, "Integrated transport strategies: a new approach to urban transport policy formulation in the UK," *Transport Reviews*, vol. 11, no. 3, pp. 223–247, 1991.
- [2] R. M. Stanger and V. R. Vuchic, "The design of bus-rail transit facilities," *Transit Journal*, vol. 5, no. 4, pp. 61–72, 1979.
- [3] J. A. Dunn Jr., "Coordination of urban transit services: the German model," *Transportation*, vol. 9, no. 1, pp. 33–43, 1980.
- [4] B. F. Byrne and V. Vuchic, "Public transportation line positions and headways for minimum cost," *Traffic Flow and Transportation*, pp. 347–360, 1972.
- [5] G. K. Kuah and J. Perl, "Optimization of feeder bus routes and bus-stop spacing," *Journal of Transportation Engineering*, vol. 114, no. 3, pp. 341–354, 1988.
- [6] G. K. Kuah and J. Perl, "The feeder-bus network-design problem," *Journal of the Operational Research Society*, vol. 40, pp. 751–767, 1989.
- [7] B. F. Byrne, "Cost minimizing positions, lengths and headways for parallel public transit lines having different speeds," *Transportation Research*, vol. 10, no. 3, pp. 209–214, 1976.
- [8] V. F. Hurdle, "Minimum cost locations for parallel public transit lines," *Transport Science*, vol. 7, no. 4, pp. 340–350, 1973.
- [9] S. C. Wirasinghe, V. F. Hurdle, and G. Newell, "Optimal parameters for a coordinated rail and bus transit system," *Transportation Science*, vol. 11, no. 4, pp. 359–374, 1977.
- [10] S. C. Wirasinghe, "Nearly optimal parameters for a rail/feeder-bus system on a rectangular grid," *Transportation Research A*, vol. 14, no. 1, pp. 33–40, 1980.
- [11] S. Chien and P. Schonfeld, "Joint optimization of a rail transit line and its feeder bus system," *Journal of Advanced Transportation*, vol. 32, no. 3, pp. 253–284, 1998.
- [12] S. Chien and Z. Yang, "Optimal feeder bus routes on irregular street networks," *Journal of Advanced Transportation*, vol. 34, no. 2, pp. 213–248, 2000.
- [13] S. Chien, Z. Yang, and E. Hou, "Genetic algorithm approach for transit route planning and design," *Journal of Transportation Engineering*, vol. 127, no. 3, pp. 200–207, 2001.
- [14] C. L. Martins and M. V. Pato, "Search strategies for the feeder bus network design problem," *European Journal of Operational Research*, vol. 106, no. 2-3, pp. 425–440, 1998.
- [15] P. Shrivastav and S. L. Dhingra, "Development of feeder routes for suburban railway stations using heuristic approach," *Journal of Transportation Engineering*, vol. 127, no. 4, pp. 334–341, 2001.
- [16] S. N. Kuan, H. L. Ong, and K. M. Ng, "Applying metaheuristics to feeder bus network design problem," *Asia-Pacific Journal of Operational Research*, vol. 21, no. 4, pp. 543–560, 2004.
- [17] S. N. Kuan, H. L. Ong, and K. M. Ng, "Solving the feeder bus network design problem by genetic algorithms and ant colony optimization," *Advances in Engineering Software*, vol. 37, no. 6, pp. 351–359, 2006.
- [18] F. Ciaffi, E. Cipriani, and M. Petrelli, "Feeder bus network design problem: a new metaheuristic procedure and real size applications," *Procedia-Social and Behavioral Sciences*, vol. 54, pp. 798–807, 2012.

## Research Article

# Study on Driving Stability of Tank Trucks Based on Equivalent Trammel Pendulum for Liquid Sloshing

Xian-sheng Li, Xue-lian Zheng, Yuan-yuan Ren, Yu-ning Wang, and Zhu-qing Cheng

*College of Traffic, Jilin University, No. 5988 Renmin Street, Changchun 130022, China*

Correspondence should be addressed to Yuan-yuan Ren; [maggie170101@gmail.com](mailto:maggie170101@gmail.com)

Received 17 July 2013; Revised 27 September 2013; Accepted 27 September 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Xian-sheng Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To investigate the driving stability of tank trucks, an equivalent trammel pendulum was utilized to approximately demonstrate the dynamic characteristics of liquid sloshing in a partially filled tank. The oscillation movement of the trammel pendulum in the tank was described under the tank-fixed coordinate system and its motion equation under the noninertia coordinate system was derived using a Lagrangian function. The motion of the pendulum that expresses the fluid cargo dynamic behavior and that of the solid truck was coupled with each other by the tank. Therefore, a tank truck dynamic model was established using Newton's first law and the angular momentum. A typical tank truck was selected and used to study its driving stability under steering angle step test. The study on tankers driving stability is of great importance for evaluating tankers driving safety, investing the main impact factor aspecting tankers driving stability, and developing active/passive roll control systems for them.

## 1. Introduction

About 80% of global chemical and petroleum products are delivered by road tank vehicles. The transportation freight has already reached 4 billion tons per year. In America, tankers make up more than 55% of all freight trucks. Tankers are hugely convenient for fluid material exchange and have a positive effect on boosting the national economic development. However, they also create severe traffic safety problems which would result in huge people injury and property damage. Statistical data collected by Statistique Canada has shown that 83% of lorry rollover accidents on highways are caused by tank vehicles [1]. Meanwhile, in 2011, 416 tanker accidents occurred in China, resulting in more than 400 people being injured or killed as well as immense economic losses. Besides this, for the particularity of liquid cargoes, the release of fluid cargo in tanker accidents could cause contamination to the roads, the water, and the air [2]. Therefore, great attention must be paid to tanker driving safety.

Much works have been carried out on the characteristics of tanker accidents, in an attempt to investigate the primary accident type. It was found by Treichel et al. that rollover is the most frequent accident type for tankers [3], comprising 45.16% of all tanker accidents in China in 2010. Furthermore,

nearly 61% of tanker rollovers occurred on curved sections of highway. Many researchers have studied the factors behind this phenomenon and have concluded that liquid sloshing in a partially filled tank is the main cause [4–7]. Due to the difference in liquid cargo densities and limitations on vehicle axle loads, tanks are in a partially filled condition most of the time. Due to the existence of liquid-free sloshing space, transient liquid sloshing is produced when vehicle driving state changes [8, 9]. The lateral sloshing force that acts on the tank wall increases vehicle rollover torque and degrades vehicle roll stability.

To explore the influence of transient liquid sloshing on tanker roll stability, many research works have been explored and some important conclusions have been drawn. Considering the complexity of tank vehicle dynamic analysis—as tank vehicles are fluid-solid coupling multibody system—some important simplifications have to be made before the analysis. Strandberg et al. assumed that liquid free surface is a tilted straight line whose gradient is a function of vehicle's lateral acceleration and tank's roll angle [10–13]. Based on that, the center of gravity (CG) of liquid bulk can be obtained, and a quasi-static (QS) dynamic analysis of vehicle driving stability can be carried out. The QS method was popular in the initial research phase of the late 1990s,

and many improved algorithms have been produced to increase the analysis precision to some degree [14, 15].

To investigate the accuracy of the QS method, some researchers studied the relationship between the evaluated liquid sloshing effect obtained by calculating the movement of the CG of the liquid bulk and the observed effect. They have found that the mean values of sloshing forces and the coordinates of the CG of the liquid bulk are quite close between the estimated and observed cases [16]. This result demonstrates the correctness of the QS method, which means that the worst vehicle driving state can be predicted by the QS method [17].

However, the QS method cannot reflect the dynamic characteristics of liquid sloshing. Therefore, an equivalent mechanical model which can precisely demonstrate the dynamic characteristic of liquid sloshing in a partially filled tank must be modeled first [18–21]. In estimating the tanker's dynamic analysis using equivalent mechanical model for liquid sloshing, the biggest difficulty for researchers is the coupling of the fluid cargo and the solid vehicle. To solve this problem, some people analyzed tanker driving stability in multibody analysis software, and some used the contractive analysis method [20, 22]. Neither method had the advantages of the equivalent mechanical model for liquid sloshing, as a vehicle's driving stability cannot be precisely described by the analytical model for vehicle roll stability control system.

Therefore, the purpose of this paper is to establish a tank vehicle dynamic model based on the equivalent mechanical model for transient liquid sloshing in a partially filled tank. The trammel pendulum was used to describe liquid sloshing, its motion was analyzed, and the motion equation was derived under the tank-fixed coordinate. Then, the coupling condition for the fluid cargo and the solid vehicle was studied, and a tanker dynamic model was established based on this. Finally, MATLAB was used to simulate the roll stability of a typical tank truck under steering angle step test, and some important conclusions were drawn. The research achievement is of great significance for commanding and evaluating tank trucks roll stability performance, judging the conditions under which a tanker will lose its driving stability and designing roll stability control systems for tankers to improve their driving safety.

## 2. Equivalent Trammel Pendulum for Liquid Sloshing in a Partially Filled Tank

*2.1. The Basic Motion Equation and Parameter Values of the Trammel Pendulum.* Theoretical analysis and experimental studies had revealed that the first-order sloshing mode, which can be described by the oscillation of liquid-free surface, is the most important mode of liquid sloshing in partially filled tanks [20, 21]. MATLAB's numerical method was used to calculate the trajectory of the CG of the liquid bulk that oscillates in a tank with an elliptical or circular cross section. The results showed that the trajectory of the CG of the liquid bulk remains parallel to the tank's periphery [21], which is presented in Figure 1. According to this fact, the trammel pendulum model whose swing trajectory is elliptic

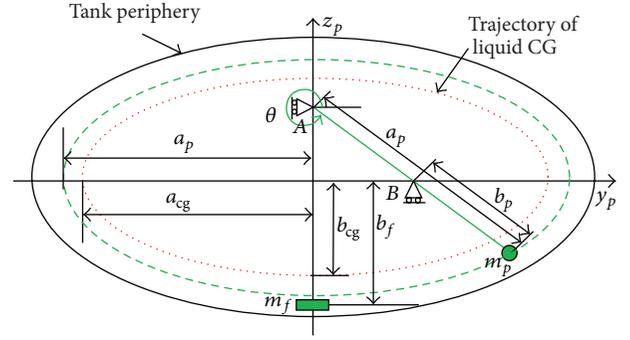


FIGURE 1: Schematic diagram for the motion of the trammel pendulum.

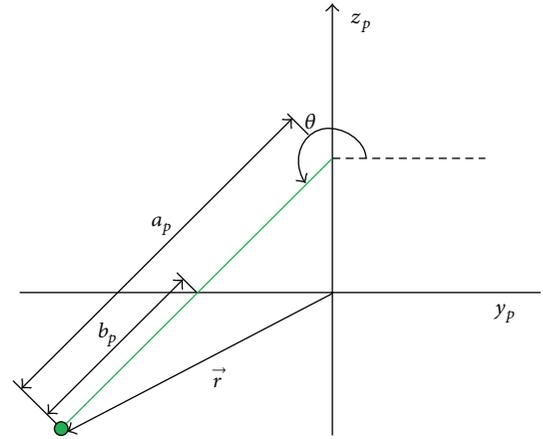


FIGURE 2: Analytical diagram for the motion of the trammel pendulum under an inertial coordinate system.

is established to approximately simulate liquid sloshing in a partially filled tank.

In Figure 1,  $a_p$  is half of the major axis of the pendulum's oscillation trajectory and  $b_p$  is half of its minor axis.  $a_{cg}$  is half of the major axis of the elliptical trajectory of the CG of the liquid bulk and  $b_{cg}$  is half of its minor axis.  $m_p$  is the pendulum mass, which is equal to the sloshing liquid mass;  $m_f$  is the fixed liquid mass, which does not participate in sloshing.  $\theta$  is the pendulum's amplitude.

The Lagrangian function was used to derive the kinetic equation of the trammel pendulum system. The motion equation of the trammel pendulum that oscillates under tank fixed inertia coordinate system whose origin locates at the center of the tank (as shown in Figure 2) is written as follows:

$$\ddot{\theta} (a_p^2 \sin^2 \theta + b_p^2 \cos^2 \theta) + \frac{1}{2} \dot{\theta}^2 (a_p^2 - b_p^2) \sin 2\theta + g b_p \cos \theta = 0. \quad (1)$$

The pendulum parameters in (1) are presented as follows:

$$\frac{b_p}{b} = 1.089 + 0.726\Delta - 0.1379\Lambda - 0.953\Delta^2 - 1.216\Lambda\Delta + 0.05141\Lambda^2 - 0.06107\Delta^3 + 0.5739\Lambda\Delta^2 + 0.1632\Lambda^2\Delta, \quad (2)$$

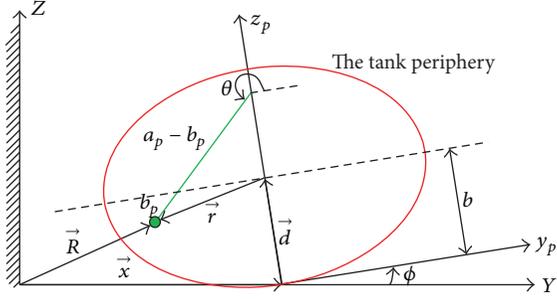


FIGURE 3: Analytical diagram for the motion of the trammel pendulum under the noninertial coordinate system.

$$\frac{m_p}{m} = 0.7844 - 1.729\Delta + 0.3351\Lambda + 1.156\Delta^2 + 0.7256\Lambda\Delta - 0.1254\Lambda^2 - 0.3219\Delta^3 - 0.9152\Lambda\Delta^2 + 0.08043\Lambda^2\Delta, \quad (3)$$

$$m_f = m - m_p, \quad (4)$$

$$\frac{(b - b_f)}{b} = \frac{[m(b - b_{cg}) - m_f(b - b_p)]}{m_f b}, \quad (5)$$

where  $\Delta$  is the liquid fill percentage, which is the ratio of the height of liquid's level surface to that of the tank, and  $\Lambda$  is the ratio of the tank's width to its height.

The detailed derivation process for (1)–(5) is given by Zheng et al. [21].

According to Figure 1, the following equation holds:

$$\frac{a}{b} = \frac{a_{cg}}{b_{cg}} = \frac{a_p}{b_p} = \Lambda. \quad (6)$$

From (6), the arm length of the pendulum can be expressed as follows:

$$a_p = b_p \times \Lambda. \quad (7)$$

As the mass of liquid bulk can easily be acquired while the tank's size and liquid density are given, the pendulum mass and the fixed liquid mass can be obtained using (3) and (4).

**2.2. Motion Equations of the Trammel Pendulum under the Noninertial Coordinate System.** During the driving process, the vehicle's driving state changes regularly under the influence of a variety of factors. It leads to the movement of the tank, and in turn the coordinate system is being fixed relative to the tank. Therefore, motion equations for a trammel pendulum under a noninertial coordinate system (translation and rotation are included) should be derived in order to study the motion characteristics of liquid bulk when tankers are being driven on curved sections of road, avoiding obstacles or changing lanes.

The schematic diagram for the motion of the tank when vehicle turns right is presented in Figure 3. For convenient analysis of the vehicle's roll stability, the origin of the tank-fixed coordinate system is located at the bottom of the tank.

This location will not impact the motion analysis result. The origin of the earth-fixed coordinate system also is located at the bottom of the tank when the vehicle is static.  $\vec{x}$  is the distance vector from the origin of the earth-fixed coordinate system to that of the tank-fixed coordinate system.  $\vec{R}$  is the absolute position vector of the pendulum mass ball.  $\vec{r}$  is the absolute position vector of the pendulum mass ball.  $\phi$  is the tank roll angle, which is defined as the angle formed by rotating the  $Y$ -axis counterclockwise till the position of the  $y_p$ -axis.

With reference to Figure 3, the absolute position of the pendulum mass ball can be obtained as follows:

$$\vec{R} = [x - b \sin \phi + (a_p \cos \theta - b_p \sin \theta \tan \phi) \cos \phi] \vec{i} + [b \cos \theta + (b_p \sin \theta + a_p \cos \theta \tan \phi) \cos \phi] \vec{j}. \quad (8)$$

The velocity of the pendulum mass ball can be obtained by solving the first-order derivative of (8), which may be presented as follows:

$$\begin{aligned} \dot{\vec{R}} = & [\dot{x} - (b \cos \phi + a_p \cos \theta \sin \phi + b_p \sin \theta \cos \phi) \dot{\phi} \\ & - (a_p \sin \theta \cos \phi + b_p \cos \theta \sin \phi) \dot{\theta}] \vec{i} \\ & + [(-b \sin \phi - b_p \sin \theta \sin \phi + a_p \cos \theta \cos \phi) \dot{\phi} \\ & + (b_p \cos \theta \cos \phi - a_p \sin \theta \sin \phi) \dot{\theta}] \vec{j}. \end{aligned} \quad (9)$$

Similarly, the acceleration of the pendulum mass ball can be acquired by solving the differential of (9), which may be written as follows:

$$\begin{aligned} \ddot{\vec{R}} = & [-\ddot{\phi} (b \cos \phi + a_p \cos \theta \sin \phi + b_p \sin \theta \cos \phi) \\ & - \ddot{\theta} (a_p \sin \theta \cos \phi + b_p \cos \theta \sin \phi) \\ & + 2\dot{\theta}\dot{\phi} (a_p \sin \theta \sin \phi - b_p \cos \theta \cos \phi) \\ & + \dot{\phi}^2 (b \sin \phi - a_p \cos \theta \cos \phi + b_p \sin \theta \sin \phi) \\ & + \dot{\theta}^2 (-a_p \cos \theta \cos \phi + b_p \sin \theta \sin \phi) + \ddot{x}] \vec{i} \\ & + [\ddot{\phi} (-b \sin \phi + a_p \cos \theta \cos \phi - b_p \sin \theta \sin \phi) \\ & + \ddot{\theta} (b_p \cos \theta \cos \phi - a_p \sin \theta \sin \phi) \\ & - \dot{\theta}^2 (a_p \cos \theta \sin \phi + b_p \sin \theta \cos \phi) \\ & + \dot{\phi}^2 (-b \cos \phi - a_p \cos \theta \sin \phi - b_p \sin \theta \cos \phi) \\ & - 2\dot{\theta}\dot{\phi} (a_p \sin \theta \cos \phi + b_p \cos \theta \sin \phi)] \vec{j}. \end{aligned} \quad (10)$$

The zero of potential energy is defined as the  $Y$ -axis, and the potential energy of the pendulum mass ball can be presented as follows:

$$V = mg(b + b_p \sin \theta + a_p \cos \theta \tan \phi) \cos \phi. \quad (11)$$

According to (10) and (11), a Lagrangian function can be used to obtain the motion equation of the pendulum system, which is written as follows:

$$L = T - V, \quad (12)$$

where  $T$  is the kinetic energy of the pendulum mass ball.

The motion of the trammel pendulum system can be expressed by

$$\frac{\partial}{\partial t} \left( \frac{\partial L}{\partial \dot{a}_j} \right) - \frac{\partial L}{\partial a_j} = 0, \quad (13)$$

where  $a_j$  is the degree of freedom for system  $L$ .

According to the motion analysis of the trammel pendulum,  $\theta$ , the pendulum amplitude, and  $\phi$ , the tank roll angle, are the two degrees of freedom for the trammel pendulum system.

The differential of (12) with respect to  $\theta$  can be expressed as follows:

$$\begin{aligned} \frac{\partial L}{\partial \theta} = m & \left[ \dot{\phi}^2 (-a_p^2 \sin \theta \cos \theta + b_p^2 \sin \theta \cos \theta + b_p b \cos \theta) \right. \\ & + \dot{\theta}^2 (a_p^2 \sin \theta \cos \theta - b_p^2 \sin \theta \cos \theta) \\ & + \dot{\theta} \dot{\phi} (a_p b \cos \theta) \\ & - \dot{x} \dot{\phi} (-a_p \sin \theta \sin \phi + b_p \cos \theta \cos \phi) \\ & - \dot{x} \dot{\theta} (a_p \cos \theta \cos \phi - b_p \sin \theta \sin \phi) \\ & \left. - g (b_p \cos \theta - a_p \sin \theta \tan \phi) \cos \phi \right]. \end{aligned} \quad (14)$$

The differential of (12) with respect to  $\dot{\theta}$  can be expressed by

$$\begin{aligned} \frac{\partial L}{\partial \dot{\theta}} = m & \left[ \dot{\theta} (a_p^2 \sin^2 \theta + b_p^2 \cos^2 \theta) + \dot{\phi} (a_p b_p + a_p b \sin \theta) \right. \\ & \left. - \dot{x} (a_p \sin \theta \cos \phi + b_p \cos \theta \sin \phi) \right]. \end{aligned} \quad (15)$$

Finally, the differential of (15) with respect to  $t$  can be expressed as follows:

$$\begin{aligned} \frac{\partial}{\partial t} \left( \frac{\partial L}{\partial \dot{\theta}} \right) = m & \left[ \ddot{\theta} (a_p^2 \sin^2 \theta + b_p^2 \cos^2 \theta) \right. \\ & + \dot{\theta}^2 (2a_p^2 \sin \theta \cos \theta - 2b_p^2 \sin \theta \cos \theta) \\ & + \ddot{\phi} (a_p b_p + a_p b \sin \theta) + \dot{\phi} \dot{\theta} a_p b \cos \theta \\ & - \ddot{x} (a_p \sin \theta \cos \phi + b_p \cos \theta \sin \phi) \\ & - \dot{x} \dot{\theta} (a_p \cos \theta \cos \phi - b_p \sin \theta \sin \phi) \\ & \left. + \dot{x} \dot{\phi} (a_p \sin \theta \sin \phi - b_p \cos \theta \cos \phi) \right]. \end{aligned} \quad (16)$$

Substituting (14)–(16) into (13), we get

$$\begin{aligned} & \ddot{\theta} (a_p^2 \sin^2 \theta + b_p^2 \cos^2 \theta) + \ddot{\phi} (a_p b_p + a_p b \sin \theta) \\ & + \dot{\phi}^2 \left[ \frac{1}{2} (a_p^2 - b_p^2) \sin 2\theta - b_p b \cos \theta \right] \\ & + \frac{1}{2} \dot{\theta}^2 (a_p^2 - b_p^2) \sin 2\theta \\ & - \ddot{x} (a_p \sin \theta \cos \phi + b_p \cos \theta \sin \phi) \\ & + g (b_p \cos \theta \cos \phi - a_p \sin \theta \sin \phi) = 0. \end{aligned} \quad (17)$$

The differential of (12) with respect to  $\phi$  can be expressed as follows:

$$\begin{aligned} \frac{\partial L}{\partial \phi} = m & \left[ \dot{x} \dot{\phi} (b \sin \phi - a_p \cos \theta \cos \phi + b_p \sin \theta \sin \phi) \right. \\ & - \dot{x} \dot{\theta} (-a_p \sin \theta \sin \phi + b_p \cos \theta \cos \phi) \\ & - g a_p \cos \theta \sec \phi \\ & \left. + g (b + b_p \sin \theta + a_p \cos \theta \tan \phi) \sin \phi \right]. \end{aligned} \quad (18)$$

The differential of (12) with respect to  $\dot{\phi}$  can be expressed by

$$\begin{aligned} \frac{\partial L}{\partial \dot{\phi}} = m & \left[ \dot{\phi} (b^2 + a_p^2 \cos^2 \theta + b_p^2 \sin^2 \theta + 2b_p b \sin \theta) \right. \\ & + \dot{\theta} (a_p b_p + a_p b \sin \theta) \\ & \left. - \dot{x} (b \cos \phi + a_p \cos \theta \sin \phi + b_p \sin \theta \cos \phi) \right]. \end{aligned} \quad (19)$$

According to (19), the differential of (19) with respect to  $t$  can be obtained as follows:

$$\begin{aligned} \frac{\partial}{\partial t} \left( \frac{\partial L}{\partial \dot{\phi}} \right) = m & \left[ \ddot{\phi} (b^2 + a_p^2 \cos^2 \theta + b_p^2 \sin^2 \theta + 2b_p b \sin \theta) \right. \\ & + \ddot{\theta} (a_p b_p + a_p b \sin \theta) + a_p b \dot{\theta}^2 \cos \theta \\ & + \dot{\phi} \dot{\theta} (-2a_p^2 \sin \theta \cos \theta + 2b_p^2 \sin \theta \cos \theta + 2b_p b \cos \theta) \\ & + \dot{x} \dot{\theta} (a_p \sin \theta \sin \phi - b_p \cos \theta \cos \phi) \\ & + \dot{x} \dot{\phi} (b \sin \phi - a_p \cos \theta \cos \phi + b_p \sin \theta \sin \phi) \\ & \left. - \dot{x} (b \cos \phi + a_p \cos \theta \sin \phi + b_p \sin \theta \cos \phi) \right]. \end{aligned} \quad (20)$$

Substituting (18)–(20) into (13) we get

$$\begin{aligned}
& \ddot{\phi} \left[ (b + b_p \sin \theta)^2 + a_p^2 \cos^2 \theta \right] \\
& + 2\dot{\theta}\dot{\phi} \left[ \frac{1}{2} \sin 2\theta (b_p^2 - a_p^2) + b_p b \cos \theta \right] \\
& + \ddot{\theta} (a_p b_p + a_p b \sin \theta) \\
& - \ddot{x} (b \cos \phi + a_p \cos \theta \sin \phi + b_p \sin \theta \cos \phi) \\
& + \dot{\theta}^2 a_p b \cos \theta \\
& - g \left[ (b + b_p \sin \theta) \sin \phi - a_p \cos \theta \cos \phi \right] = 0.
\end{aligned} \tag{21}$$

Equations (17) and (21) are the equations of motion of the trammel pendulum under the noninertia coordinate system, and the pendulum motion characteristics under the translation and rotation of the coordinate system can be analyzed using (17) and (21).

### 3. Tank Truck Dynamic Model

**3.1. Assumptions.** The roll stability of a tank truck is greatly influenced by transient liquid sloshing in partially filled tanks. Due to the flow characteristic of liquid bulk, transient liquid sloshing is produced when the vehicle's driving state changes, which leads to the oscillation of the pendulum mass ball. To simplify the analysis of the tank truck's roll stability, some assumptions are made and listed as follows.

- (1) The sprung mass is the vehicle mass that lies over the suspension system, in which the mass of the liquid cargo is not included. In this paper, vehicle and liquid bulk, as the solid and fluid parts respectively, are not mixed with each other.
- (2) In the transverse direction, the CG of the sprung mass is located at the bottom of the tank. Moreover, the sprung mass is symmetrical about the longitudinal axis of the vehicle.
- (3) Transient liquid sloshing is merely produced along the transverse direction; longitudinal liquid sloshing is not taken into consideration. The movement and mass distribution of the liquid bulk are completely the same at different tank cross sections, which means that the CG of liquid bulk is located in the middle of the tank along the longitudinal direction.
- (4) The liquid-free surface is not broken during the vehicle's driving process.

**3.2. Force Analysis of the Tank Truck.** Based on the assumptions given in Section 3.1, a reference coordinate system for tank truck dynamic analysis was selected first.

The coordinate system being fixed relative to the vehicle was chosen as the reference coordinate system for the dynamic analysis of the tank truck and the establishment of the vehicle's dynamic model. The origin of the coordinate system is the point where the vertical line that goes through

the CG of the vehicle when it is static intersects with the vehicle's roll axis.

The coordinate system being fixed relative to the vehicle's sprung mass was defined as the  $x$ - $y$ - $z$  system, where the  $x$ -axis is parallel to the longitudinal axis of the vehicle and points in the direction in which the vehicle is being driven, the  $y$ -axis is perpendicular to the  $x$ -axis and points to the left from the driver's perspective, and the  $z$ -axis is perpendicular to the  $xy$  plane and points straight up.

The vehicle's unsprung mass is assumed not to produce roll movement.

The top view and the back view of the vehicle's dynamic analysis are plotted in Figures 4 and 5. According to Figure 5, the lateral inertial forces act on the sprung mass, the unsprung mass, the fixed liquid mass, and the pendulum mass, making up the vehicle's lateral inertial force. The tire cornering forces make up the external lateral force that acts on the vehicle according to Figure 4. On the basis of Newton's first law, when the vehicle is steady, its inertial lateral force and external lateral force are equal in value and positive in direction, which can be expressed as follows:

$$(m_t + m_f) a_s + m_p a_f + m_u a_u = 2(F_f + F_r), \tag{22}$$

where  $a_s$  is the lateral acceleration of sprung mass;  $a_u$  is the lateral acceleration of unsprung mass;  $a_f$  is the lateral acceleration of the pendulum mass ball.  $m_t$  is sprung mass;  $m_u$  is unsprung mass.  $F_f$  is the front tire cornering force and  $F_r$  is the rear tire cornering force.

Generally, the fixed liquid mass does not stay static but slides slowly with the aid of gravity while the tank tilts. However, the shift of the fixed liquid mass is quite small compared to that of the pendulum mass ball and its shift amount drops greatly with the increase in the liquid fill percentage. Therefore, the fixed liquid mass is assumed to be static and its position is assumed to stay constant as the tank rolls over. Based on this assumption, the fixed liquid mass can be seen as the solid part and its lateral acceleration is equal to that of the vehicle's sprung mass.

According to the dynamic analysis of the vehicle, the lateral acceleration of the sprung mass and unsprung mass can be presented by

$$\begin{aligned}
a_s &= V(\dot{\beta} + r) - h_s \ddot{\phi} + c\dot{r}, \\
a_u &= V(\dot{\beta} + r) - e\dot{r},
\end{aligned} \tag{23}$$

where  $V$  is the vehicle's driving speed,  $\beta$  the vehicle's slip angle,  $r$  the vehicle's yaw angle,  $h_s$  is the vertical distance from the CG of sprung mass to the roll center,  $c$  is the longitudinal distance between the CG of sprung mass and that of the tank truck, and  $e$  is the longitudinal distance between the CG of unsprung mass and that of the tank truck.

According to (10), the lateral acceleration of the pendulum mass ball can be written as follows:

$$a_f = V\dot{\phi} - D_1\ddot{\phi} - D_2\ddot{\theta} + 2D_3\dot{\phi}\dot{\theta} + D_4\dot{\phi}^2 + D_5\dot{\theta}^2, \tag{24}$$

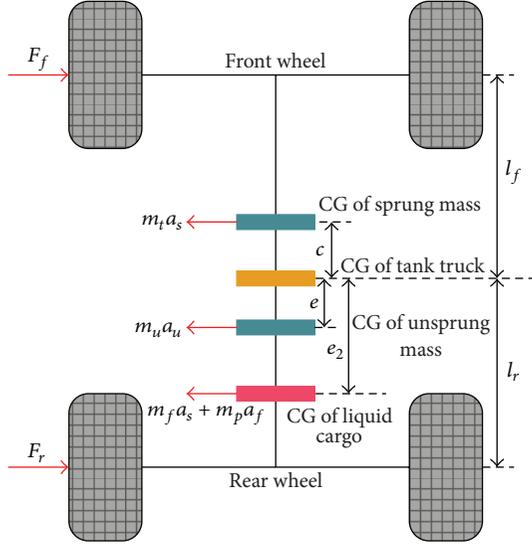


FIGURE 4: Top view of vehicle dynamic analysis.

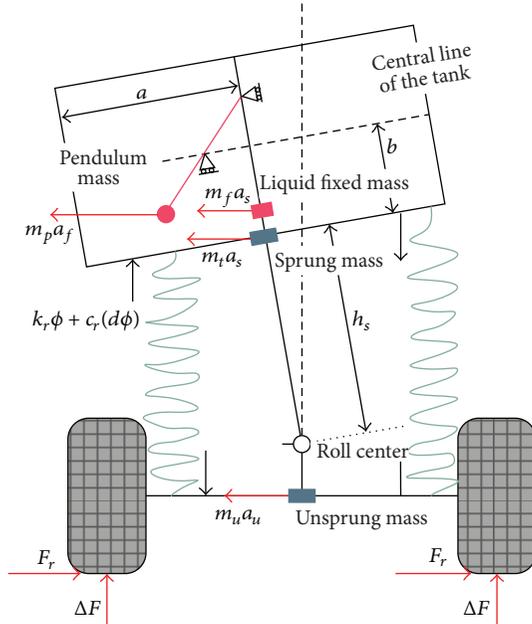


FIGURE 5: Back view of vehicle dynamic analysis.

where

$$\begin{aligned}
 D_1 &= b \cos \phi + a_p \cos \theta \sin \phi + b_p \sin \theta \cos \phi, \\
 D_2 &= a_p \sin \theta \cos \phi + b_p \cos \theta \sin \phi, \\
 D_3 &= a_p \sin \theta \sin \phi - b_p \cos \theta \cos \phi, \\
 D_4 &= b \sin \phi - a_p \cos \theta \cos \phi + b_p \sin \theta \sin \phi, \\
 D_5 &= -a_p \cos \theta \cos \phi + b_p \sin \theta \sin \phi.
 \end{aligned} \tag{25}$$

The force analysis in the roll plane is plotted in Figure 5. Due to the lateral inertia force, the sprung mass and the liquid

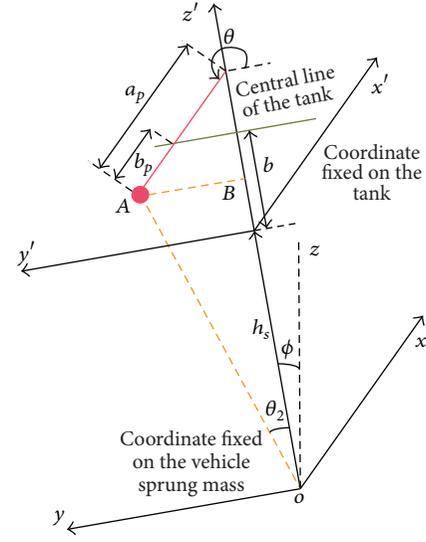


FIGURE 6: The schematic diagram for the pendulum rotating around the roll axis.

bulk rotate about the roll axis and bring about rollover torque. The roll angle about the roll axis for sprung mass and the fixed liquid mass is \$\phi\$, which can be seen from Figure 5. However, owing to the tank and the pendulum mass ball movement, the roll angle about the roll axis for the pendulum mass ball is composed of two parts, which is shown in Figure 6.

According to Figure 6, the roll angle about the roll axis for the pendulum mass ball can be expressed as follows:

$$\angle = \phi + \theta_2, \tag{26}$$

where

$$\theta_2 = \arctan \left( \frac{AB}{BO} \right) = \arctan \left( \frac{-a_p \cos \theta}{h_s + b + b_p \sin \theta} \right). \tag{27}$$

Therefore, the absolute roll rate and roll angular acceleration can be obtained as follows based on (26):

$$\dot{\angle} = \dot{\phi} + \dot{\theta}_2; \quad \ddot{\angle} = \ddot{\phi} + \ddot{\theta}_2, \tag{28}$$

$$\dot{\theta}_2 = \frac{\dot{\theta} B_1}{B}; \quad \ddot{\theta}_2 = \frac{(A_1 \ddot{\theta} - A_2 \dot{\theta}^2)}{A}, \tag{29}$$

where

$$\begin{aligned}
 B &= (h_s + b + b_p \sin \theta)^2 + (a_p \cos \theta)^2; \\
 B_1 &= a_p [(h_s + b) \sin \theta + b_p]; \\
 A &= B^2; \quad A_1 = a_p [(h_s + b) \sin \theta + b_p] B; \\
 A_2 &= 2a_p \cos \theta [(h_s + b) \sin \theta + b_p] \\
 &\quad \times [b_p (h_s + b + b_p \sin \theta) - a_p^2 \sin \theta] \\
 &\quad - a_p B (h_s + b) \cos \theta.
 \end{aligned} \tag{30}$$

On the basis of (28), the vehicle's roll balance equation and yaw balance equation can be acquired by differentiating the angular momentum with respect to time.

In general, the angular momentum of a rotating object is defined as the product of a body's rotational inertia tensor and rotational velocities about a particular coordinate.

(1) *The Angular Momentum of Sprung Mass and the Fixed Liquid Mass.* As the vehicle sprung mass and the fixed liquid mass have the same rotational velocity about the  $x$ - $y$ - $z$  coordinate system, the angular momentum for the two parts is solved together.

According to the definition of angular momentum, that for the sprung mass and fixed liquid mass can be expressed as follows:

$$H_s = \begin{bmatrix} I_{xxs} + I_{xxf} & -I_{xys} - I_{xyf} & -I_{xzs} - I_{xzf} \\ -I_{yxs} - I_{yxf} & I_{yys} + I_{yyf} & -I_{yzs} - I_{yzf} \\ -I_{zxs} - I_{zxf} & -I_{zys} - I_{zyf} & I_{zzs} + I_{zzf} \end{bmatrix} \begin{bmatrix} \dot{\phi} \\ 0 \\ r \end{bmatrix}, \quad (31)$$

where  $I_{xxs}$ ,  $I_{yys}$ , and  $I_{zzs}$  are the moments of inertia for the sprung mass about the  $x$ -axis,  $y$ -axis, and  $z$ -axis, respectively;  $I_{xxf}$ ,  $I_{yyf}$ , and  $I_{zzf}$  are the moments of inertia for the fixed liquid mass about the  $x$ -axis,  $y$ -axis, and  $z$ -axis, respectively;  $I_{xys} = I_{yxs}$ ,  $I_{xzs} = I_{zxs}$ , and  $I_{yzs} = I_{zys}$  are the products of inertia for the sprung mass about the  $x$ - and  $y$ -axis,  $x$ - and  $z$ -axis, and  $y$ - and  $z$ -axis, respectively;  $I_{xyf} = I_{yxf}$ ,  $I_{xzf} = I_{zxf}$ , and  $I_{yzf} = I_{zyf}$  are the products of inertia for the fixed liquid mass about the  $x$ - and  $y$ -axis,  $x$ - and  $z$ -axis, and  $y$ - and  $z$ -axis, respectively.

According to the assumptions given in Section 3.1, the distribution of the vehicle's sprung mass is symmetrical about the  $xz$ -plane. Therefore, the products of inertia for the sprung mass can be expressed as follows:

$$I_{xys} = 0; \quad I_{zys} = 0. \quad (32)$$

As the fixed liquid mass does not move while the tank tilts, the distribution of the fixed liquid mass is symmetrical about the  $x_t z_t$ -plane and the  $xz$ -plane. Therefore, the products of inertia for the fixed liquid mass can be expressed as follows:

$$I_{xyf} = 0; \quad I_{zyf} = 0. \quad (33)$$

Therefore, based on (31)–(33), the angular momentum for the vehicle's sprung mass and the fixed liquid mass can be written as follows:

$$H_s = \left[ (I_{xxs} + I_{xxf}) \dot{\phi} - (I_{xzs} + I_{xzf}) r \right] \vec{i} + \left[ -(I_{zxs} + I_{zxf}) \dot{\phi} + (I_{zzs} + I_{zzf}) r \right] \vec{j}. \quad (34)$$

(2) *The Angular Momentum of the Pendulum Mass.* According to (28), the angular momentum of the pendulum mass can be expressed as follows:

$$H_p = \begin{bmatrix} I_{xxp} & -I_{xyp} & -I_{xzp} \\ -I_{yxp} & I_{yyp} & -I_{yzp} \\ -I_{zxp} & -I_{zyp} & I_{zyp} \end{bmatrix} \begin{bmatrix} \dot{\phi} + \dot{\theta}_2 \\ 0 \\ r \end{bmatrix}, \quad (35)$$

where  $I_{xxp}$ ,  $I_{yyp}$ , and  $I_{zyp}$  are the moments of inertia for the pendulum mass about the  $x$ -axis,  $y$ -axis, and  $z$ -axis,

respectively;  $I_{xyp} = I_{yxp}$ ,  $I_{xzp} = I_{zxp}$ , and  $I_{yzp} = I_{zyp}$  are the products of inertia for the pendulum mass about the  $x$ - and  $y$ -axis, the  $x$ - and  $z$ -axis, and the  $y$ - and  $z$ -axis, respectively.

The expanded form of (35) is written as

$$H_p = \left[ I_{xxp} (\dot{\phi} + \dot{\theta}_2) - I_{xzp} r \right] \vec{i} + \left[ -I_{xyp} (\dot{\phi} + \dot{\theta}_2) - I_{yzp} r \right] \vec{j} + \left[ -I_{xzp} (\dot{\phi} + \dot{\theta}_2) + I_{zyp} r \right] \vec{k}. \quad (36)$$

(3) *The Angular Momentum of the Unsprung Mass.* The yaw movement is the only rotational degree of freedom for the unsprung mass. Therefore, the angular momentum of the unsprung mass can be obtained by multiplying together the moment of inertia about the  $z$ -axis and the yaw rate about that, which is written as follows:

$$H_u = I_{zzu} r \vec{k}, \quad (37)$$

where  $I_{zzu}$  is the moment of inertia about the  $z$ -axis for the unsprung mass.

The change in the vehicle fixed coordinate after it has been driving for time  $\Delta t$  is plotted in Figure 7, from which we can see that the differentials of the  $\vec{i}$ ,  $\vec{j}$ , and  $\vec{k}$  unit vectors can be expressed as follows:

$$\dot{\vec{i}} = r \vec{j}; \quad \dot{\vec{j}} = -r \vec{i} + \dot{\phi} \vec{k}; \quad \dot{\vec{k}} = -\dot{\phi} \vec{j}. \quad (38)$$

The first-order derivatives of (34), (36), and (37) can be obtained using (38). Thus, the vehicle's roll moment and yaw moment have been acquired.

According to the moment balance between the vehicle inertia moment and the external moment, the vehicle's roll and yaw moment balance equation can be expressed as follows:

$$\begin{aligned} I_z \dot{r} - I_{xz} \ddot{\phi} - m_f a_s e_2 - m_p a_f e_2 - I_{xzp} (\ddot{\phi} + \ddot{\theta}) + I_{zyp} \dot{r} \\ - I_{xyp} (\dot{\phi} + \dot{\theta})^2 - I_{yzp} r (\dot{\phi} + \dot{\theta}) = 2 (F_f l_f - F_r l_r), \\ I_x \ddot{\phi} - I_{xz} \dot{r} + m_t h_s V (\dot{\beta} + r) + H_1 m_f a_s + (h_s + b) m_p a_f \\ + I_{xxp} (\ddot{\phi} + \ddot{\theta}) - I_{xzp} \dot{r} + I_{xyp} r (\dot{\phi} + \dot{\theta}) + I_{yzp} r^2 \\ = -k_\phi \phi - c_\phi \dot{\phi} \\ + \phi (m_t g h_s + m_f g H_1 + m_p g (h_s + b)) - m_p g a_p \cos \theta, \end{aligned} \quad (39)$$

where  $l_f$  is the longitudinal distance from the CG of the tank truck to the front wheel,  $l_r$  is the longitudinal distance from the CG of the tank truck to the rear wheel,  $k_\phi$  is the suspension roll stiffness,  $c_\phi$  is the suspension roll damping,  $e_2$  is the longitudinal distance from the CG of the tank truck to the CG of liquid cargo, and  $H_1$  is the vertical distance from the CG of the fixed liquid mass to the roll center.

Based on the parallel axis theorem of the moment of inertia, the following equation can be obtained:

$$\begin{aligned} I_z &= I_{zzs} + I_{zzu} + m_t c^2 + m_u e^2 + I_{zzf}, \\ I_x &= I_{xzs} + I_{xxf} + m_t h_s^2, \\ I_{zx} &= I_{xzs} + I_{zzf} + m_t h_s c. \end{aligned} \quad (40)$$

The dynamic model of the tank truck is composed of (17) and (39)–(40).

The tire cornering force in the dynamic model can be obtained using the magic formula for tires, which is given by

$$F = D \sin(C \arctan(B\alpha - E(B\alpha - \arctan B\alpha))), \quad (41)$$

where  $\alpha$  is the tire's sideslip angle, and  $B$ ,  $C$ ,  $D$ , and  $E$  are the constant parameters of the magic formula for tires.

The tire sideslip angle in (41) can be expressed as follows:

$$\begin{aligned} \alpha_f &= \arctan\left(\frac{V\beta + r l_f}{V}\right) - \delta_f, \\ \alpha_r &= \arctan\left(\frac{V\beta - r l_r}{V}\right) - \delta_r, \end{aligned} \quad (42)$$

where  $\alpha_f$  is the front tire sideslip angle and  $\alpha_r$  is the rear tire sideslip angle;  $\delta_f$  is the front wheel steering angle and  $\delta_r$  is the rear wheel steering angle.

For the convenience of vehicle dynamic simulation, the tank truck dynamic model is translated into the following form:

$$Mx = N, \quad (43)$$

and (43) can be rewritten as follows:

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \begin{bmatrix} \dot{\beta} \\ \dot{r} \\ \ddot{\phi} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \end{bmatrix}, \quad (44)$$

where

$$\begin{aligned} m_{11} &= V(m_t + m_f + m_p + m_u), \\ m_{12} &= c(m_t + m_f) - em_u, \\ m_{13} &= -h_s(m_t + m_f) - m_p D_1, \\ m_{14} &= -m_p D_2, \\ m_{21} &= V e_2(-m_f - m_p), \\ m_{22} &= I_z - m_f c e_2 + I_{zzp}, \end{aligned}$$

$$m_{23} = -I_{xz} + m_f h_s e_2 + m_p D_1 e_2 - I_{xzp},$$

$$m_{24} = m_p D_2 e_2 - \frac{I_{xzp} A_1}{A},$$

$$m_{31} = V(m_t h_s + m_f H_1 + m_p H),$$

$$m_{32} = -I_{xz} + m_f c H_1 - I_{xzp},$$

$$m_{33} = I_x - m_f h_s H_1 - m_p H D_1 + I_{xyp},$$

$$m_{34} = \frac{I_{xyp} A_1}{A} - m_p H D_2,$$

$$m_{41} = V(a_p \sin \theta \cos \phi + b_p \cos \theta \sin \phi),$$

$$m_{42} = 0,$$

$$m_{43} = a_p b_p + a_p b \sin \theta,$$

$$m_{44} = a_p^2 \sin^2 \theta + b_p^2 \cos^2 \theta,$$

$$\begin{aligned} n_1 &= 2(F_f + F_r) - rV(m_t + m_f + m_u) \\ &\quad - m_p(2D_3 \dot{\theta} \dot{\phi} + D_4 \dot{\phi}^2 + D_5 \dot{\theta}^2), \end{aligned}$$

$$\begin{aligned} n_2 &= 2(F_f l_f - F_r l_r) + m_f e_2 V r \\ &\quad + m_p e_2(2D_3 \dot{\theta} \dot{\phi} + D_4 \dot{\phi}^2 + D_5 \dot{\theta}^2) \\ &\quad - \frac{I_{xzp} \dot{\theta}^2 A_2}{A} + I_{xyp} \left( \dot{\phi} + \frac{\dot{\theta} B_1}{B} \right)^2 \\ &\quad + I_{yzp} r \left( \dot{\phi} + \frac{\dot{\theta} B_1}{B} \right), \end{aligned}$$

$$\begin{aligned} n_3 &= -c_\phi \dot{\phi} + \phi(-k_\phi + m_t g h_s + m_f g H_1 + m_p g H) \\ &\quad - a_p m_p g \cos \theta - V r(m_t h_s + m_f H_1) \\ &\quad - H m_p(2D_3 \dot{\theta} \dot{\phi} + D_4 \dot{\phi}^2 + D_5 \dot{\theta}^2) \end{aligned}$$

$$- I_{xyp} r(\dot{\phi} + \dot{\theta}) - r^2 I_{yzp} + \frac{\dot{\theta}^2 I_{xyp} A_2}{A},$$

$$n_4 = -g(b_p \cos \theta \cos \phi - a_p \sin \theta \sin \phi)$$

$$- \frac{1}{2} \dot{\theta}^2 (a_p^2 - b_p^2) \sin 2\theta$$

$$- \dot{\phi}^2 \left[ \frac{1}{2} (a_p^2 - b_p^2) \sin 2\theta - b_p b \cos \theta \right]. \quad (45)$$

According to the Cramer rule, the variables in (43) can be expressed as follows while the matrix  $M$  has a nonzero determinant:

$$\begin{aligned} x_1 &= \frac{M_1}{N}, & x_2 &= \frac{M_2}{N}, \\ x_3 &= \frac{M_3}{N}, & x_4 &= \frac{M_4}{N}, \end{aligned} \quad (46)$$

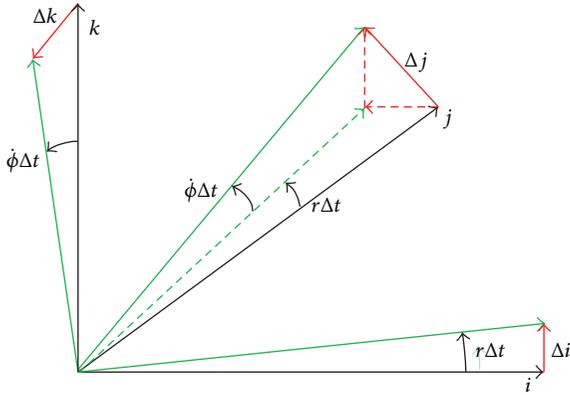


FIGURE 7: The change of the  $x$ - $y$ - $z$  coordinate system.

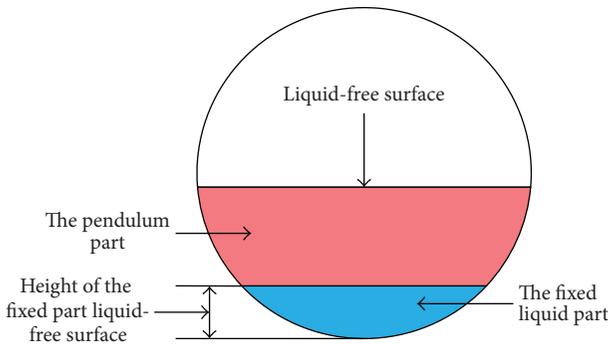


FIGURE 8: Schematic diagram for the division into the pendulum part and the fixed liquid part.

where  $M_i$ ,  $i = 1, 2, 3, 4$  is the matrix formed by replacing the  $i$ th column of  $M$  by the column vector  $N$ .

Based on (46), MATLAB's ODE algorithm is used to solve the differential equations of the vehicle's dynamic model.

**3.3. The Inertia Tensor of the Pendulum Mass and the Fixed Liquid Mass.** Before solving the differential equations of the vehicle's dynamic model, the inertia tensor for the fixed liquid mass and the pendulum mass are needed.

It is known that the liquid-free surface can be approximately described by a tilted straight line when transient liquid sloshing occurs. This means that the inertia tensor for the liquid bulk can be obtained. However, the obtained values are the sum of the pendulum part and the fixed liquid part. As it is quite difficult to divide liquid bulk physically into the pendulum part and the fixed liquid part, the fixed liquid part is assumed to locate at the bottom of the liquid bulk (as shown in Figure 8) and the height of the fixed liquid-free surface can be acquired using (3) and (4), while the density of the liquid cargo is already given. Based on this assumption, the inertia tensor for the fixed liquid part can be easily obtained. Hence, the inertia tensor for the pendulum part can also be obtained.

According to a market survey, the cross-sectional area of cylindrical tanks is usually close to  $2.4 \text{ m}^2$ . *HT5250GYQ3C*, a typical tank truck of the *HongTu* brand, is chosen as the simulation object in this paper. The radius of the tank

cross section is 1.8 m and the thickness of the tank walls is neglected. The length of the tank is 9 m. The liquid cargo is assumed to be water, whose density is  $1000 \text{ kg/m}^3$ . CAD is used to calculate the inertia tensors for the pendulum mass and the fixed liquid mass as the liquid fill level changes from 0.1 to 0.9 with a 0.1 step size and as the tilt angle of the liquid-free surface changes from 0 degrees to 90 degrees with a 10-degree step size. The calculation results for liquid fill levels of 0.2, 0.5, and 0.8 are given in Tables 1, 2, and 3.

As the fixed liquid part is assumed to be static as the tank tilts, the inertia tensor of the fixed liquid mass is always constant.

Polynomials are fitted for the data points listed in Tables 1, 2, and 3, respectively, to obtain equations that describe the inertia tensor as a function of the liquid fill percentage and the liquid-free surface tilt angle. For arbitrary tilt angle of liquid-free surface under constant liquid fill level, interpolation is used to obtain the inertia tensor of the pendulum mass.

## 4. Results and Discussion

**4.1. Tank Truck's Driving Stability When the Liquid Fill Level Is 0.5.** A MATLAB simulation is carried out for the steering angle step test of the *HongTu* tank truck with a driving speed of 90 km/h and a liquid fill level of 0.5. The simulation results for different driving variables are plotted in Figure 9. Simulation results for a normal truck that has the same parameter values, load situation (i.e., 50% laden), and driving velocity are plotted in Figure 10.

It is quite obvious that the vehicle's roll stability is greatly influenced by transient liquid sloshing in a partially filled tank. While the curves of the driving variables for the normal truck quickly return to the steady state, those for the tank truck fluctuate up and down for about 40 s before becoming steady, and the overshoot is quite large. However, the overshoot dampens quickly with the passage of time, and the damping frequency is close to the pendulum oscillation frequency. It can also be seen that the bigger the steering angle is, the more significant the effect on vehicle roll stability given by transient liquid sloshing is. As can be seen in Figure 9(d), the maximum roll rate of the tank truck is 0.2 rad/s while the steering angle is 0.02 rad, this value increases to 0.7 rad/s when the steering angle is 0.05 rad.

When the steering angle is 0.05 rad, the pendulum's amplitude changes from its original value of 4.71 rad (270 degrees) to 5.09 rad (291.7 degrees) and stays steady with the aid of lateral acceleration and gravity. The movement of the pendulum mass ball increases the vehicle's rollover torque and results in a bigger roll angle for the tank truck (Figure 9(c)) compared to the normal truck (Figure 10(c)).

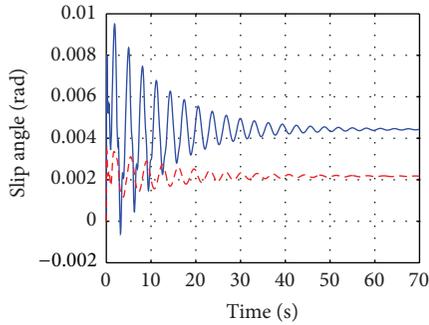
As the maximum roll rate of the vehicle is already quite large even though the steering angle is small, it seems that the vehicle may encounter rollover. However, as seen in Figure 9, the tank truck does not experience rollover but returns to a steady state after about 50 s of fluctuation, which means that whether or not the vehicle loses roll stability mostly depends on the duration of its transient response.

TABLE 1: The inertia tensors of the pendulum mass and the fixed liquid mass when the liquid fill level is 0.2.

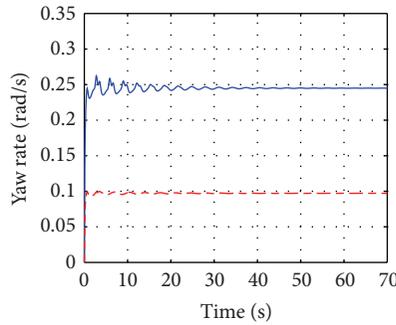
	$I_{xyp}$	$I_{zyp}$	$I_{xyp}$	$I_{xzp}$	$I_{yzp}$	$I_{xxf}$	$I_{zzf}$	$I_{xxf}$
0°	1831	18574	0	597	0	215	4088	101
10°	1926	18613	117	607	338	215	4088	101
20°	2208	18711	230	637	684	215	4088	101
30°	2670	18871	336	687	1044	215	4088	101
40°	3298	19075	432	754	1420	215	4088	101
50°	4071	19283	515	837	1806	215	4088	101
60°	4967	19485	582	933	2192	215	4088	101
70°	5959	19650	632	1039	2558	215	4088	101
80°	7020	19770	662	1153	2883	215	4088	101
90°	8101	19783	672	1268	3135	215	4088	101

TABLE 2: The inertia tensors of the pendulum mass and the fixed liquid mass when the liquid fill level is 0.5.

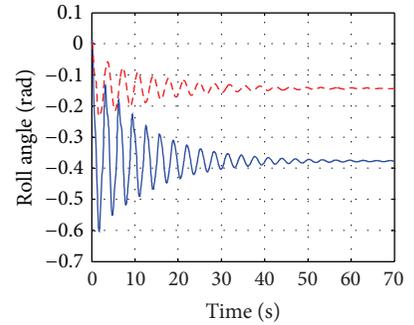
	$I_{xyp}$	$I_{zyp}$	$I_{xyp}$	$I_{xzp}$	$I_{yzp}$	$I_{xxf}$	$I_{zzf}$	$I_{xxf}$
0°	10837	45002	0	2285	0	3998	35642	1212
10°	11023	45002	228	2305	1063	3998	35642	1212
20°	11575	45002	449	2364	2094	3998	35642	1212
30°	12477	45002	656	2461	3062	3998	35642	1212
40°	13702	45002	843	2592	3936	3998	35642	1212
50°	15212	45002	1005	2754	4691	3998	35642	1212
60°	16960	45002	1136	2941	5303	3998	35642	1212
70°	18895	45002	1233	3149	5754	3998	35642	1212
80°	20957	45002	1292	3370	6031	3998	35642	1212
90°	23084	45002	1312	3598	6124	3998	35642	1212



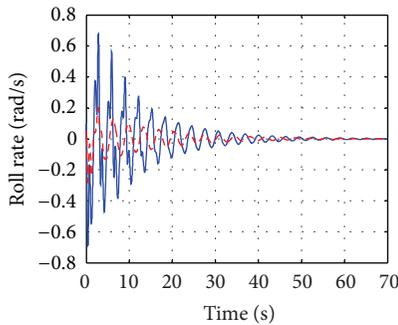
(a) Curves of slip angle



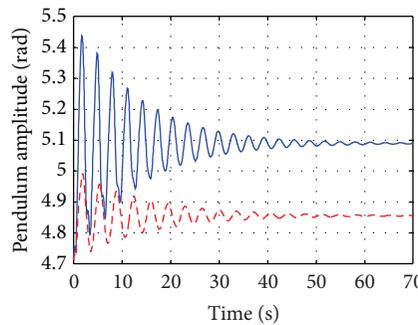
(b) Curves of yaw rate



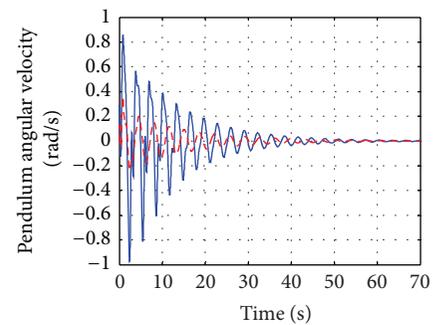
(c) Curves of roll angle



(d) Curves of roll rate



(e) Curves of pendulum amplitude



(f) Curves of pendulum angular velocity

FIGURE 9: Truck driving stability when liquid fill level is 0.5.

TABLE 3: The inertia tensors of the pendulum mass and the fixed liquid mass when the liquid fill level is 0.8.

	$I_{xpp}$	$I_{zpp}$	$I_{xyp}$	$I_{xzp}$	$I_{yzp}$	$I_{xxf}$	$I_{zzf}$	$I_{zxf}$
0°	16269	34190	0	2570	0	23308	104440	5008
10°	16363	34150	117	2579	751	23308	104440	5008
20°	16644	34040	230	2610	1461	23308	104440	5008
30°	17107	33880	336	2659	2091	23308	104440	5008
40°	17732	33680	432	2726	2611	23308	104440	5008
50°	18506	33470	515	2809	2998	23308	104440	5008
60°	19400	33270	582	2905	3240	23308	104440	5008
70°	20387	33100	632	3010	3337	23308	104440	5008
80°	21438	32980	662	3122	3298	23308	104440	5008
90°	22539	32980	672	3241	3135	23308	104440	5008

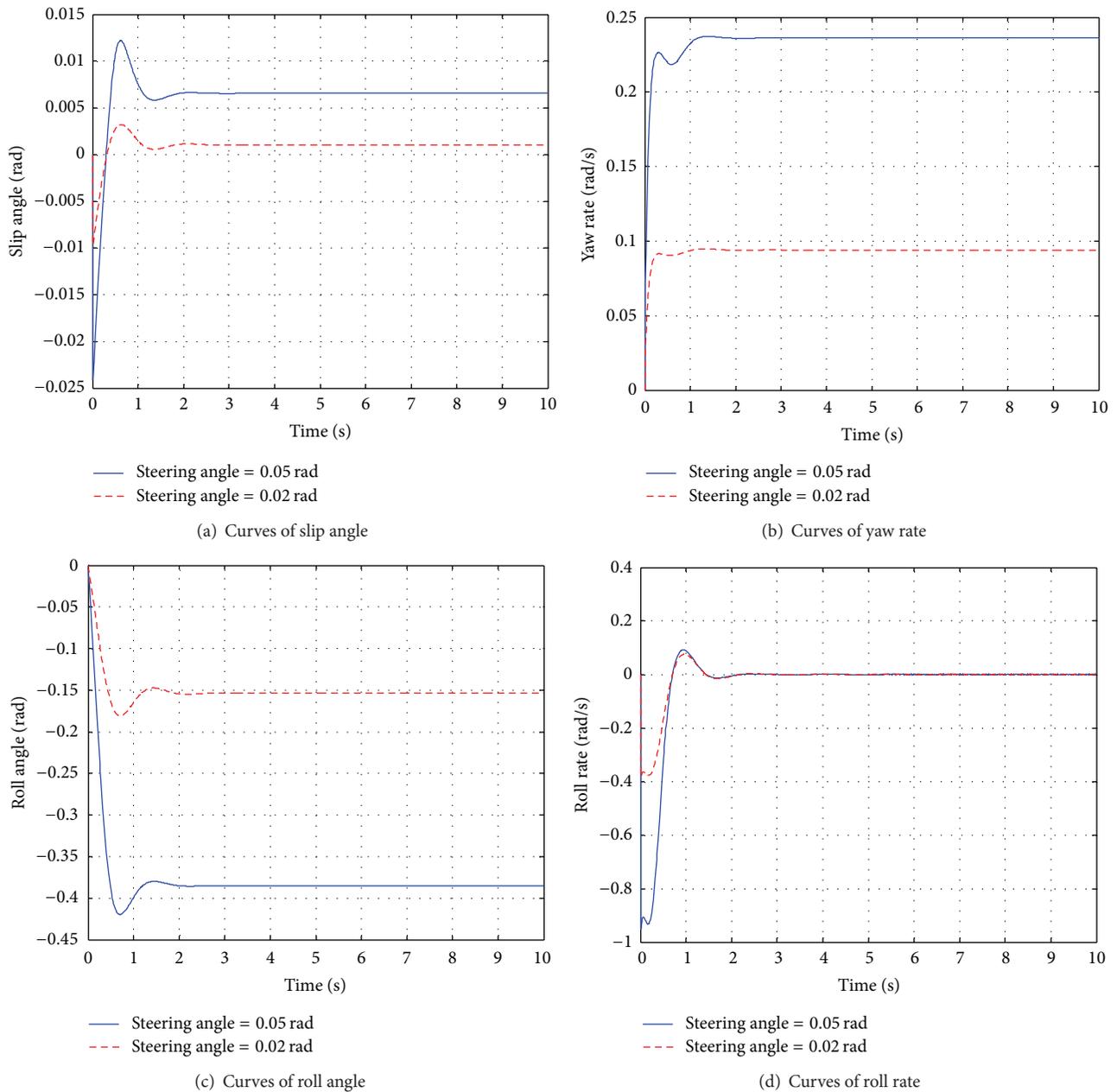


FIGURE 10: Driving stability of normal truck with 50% laden situation.

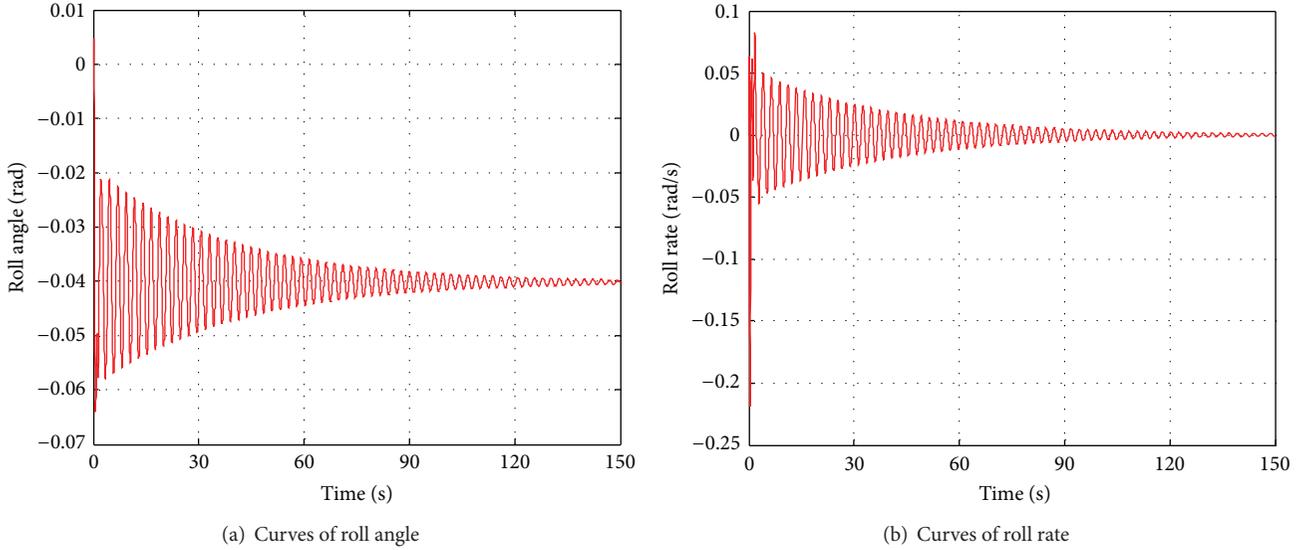


FIGURE 11: Tank truck driving stability when liquid fill level is 0.2.

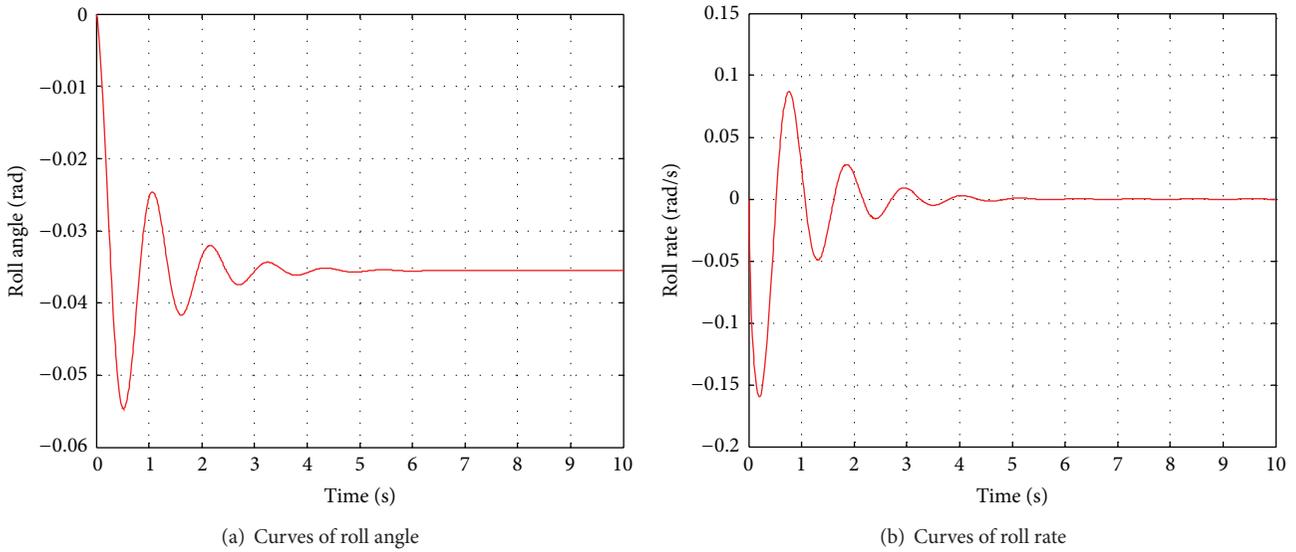


FIGURE 12: Driving stability of normal truck that is 20% laden.

From the comparison of Figures 9 and 10, it can also be seen that, although the overshoots of the driving variables of the tank truck are much bigger than those of the normal truck, the difference between their steady values is not so big, which means that the transient response of the tank truck is greatly affected by liquid sloshing, while the steady response is only slightly influenced by it.

*4.2. Tank Truck's Driving Stability When the Liquid Fill Levels Are 0.2 and 0.8.* To investigate the impact of the liquid fill level on the vehicle's roll stability, tank vehicles with liquid fill levels of 0.2 and 0.8, with a steering angle of 0.02 rad, and with a driving speed of 90 km/h are studied in a steering angle step

test. The simulation results for the roll angle and roll rate are plotted in Figures 11 and 13.

At the same time, the roll stability of normal trucks with the same steering angle, parameter values, load situations, and driving speed are also investigated to determine the influence of transient liquid sloshing in a partially filled tank on vehicle roll stability. The simulation results for the roll angle and roll rate for the normal trucks are presented in Figures 12 and 14.

As can be seen from Figure 11, when the liquid fill level is 0.2, the vehicle takes about 120 s to come to a steady state. The fluctuation times of the vehicle driving variables are much bigger than that when the liquid fill level was 0.5, as can be seen by comparing Figures 9 and 11. However,

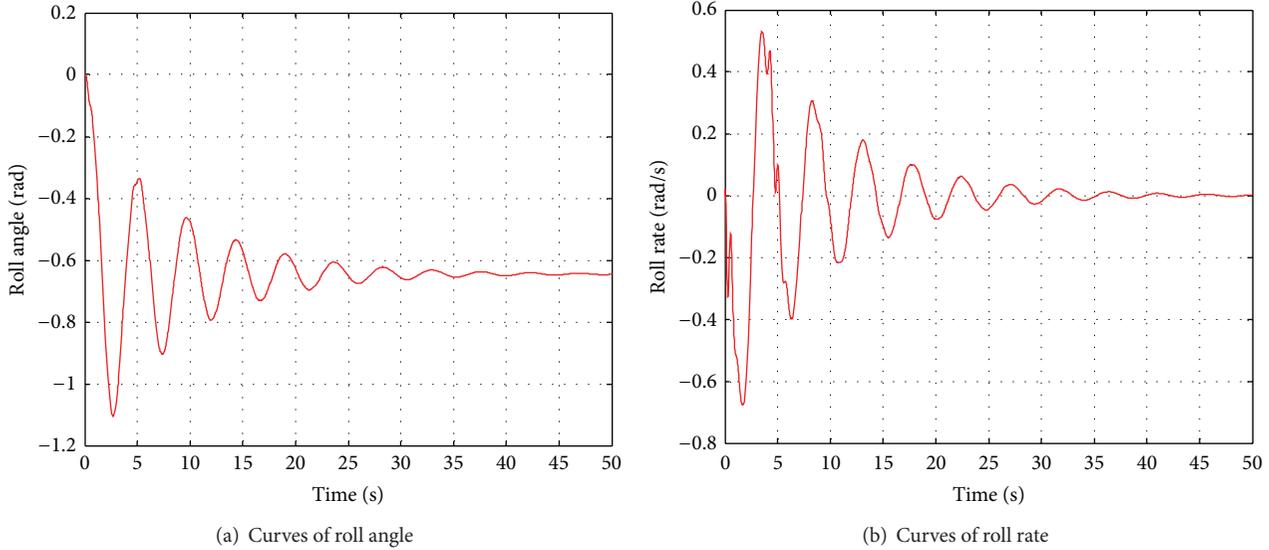


FIGURE 13: Tank truck driving stability when liquid fill level is 0.8.

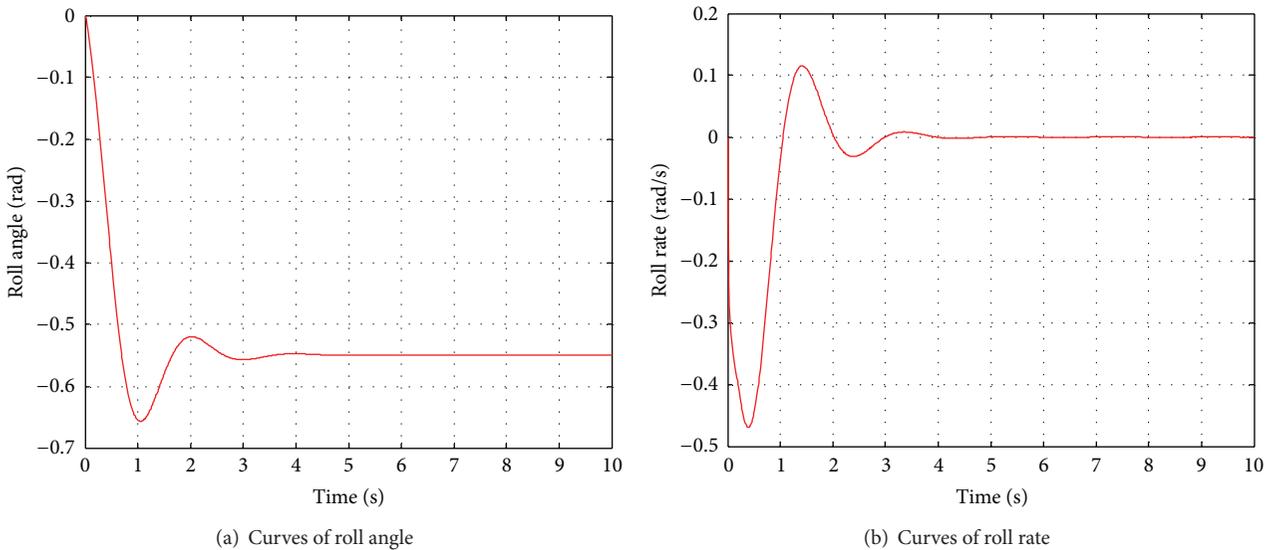


FIGURE 14: Driving stability of normal truck that is 80% laden.

the oscillation amplitude is reduced when the liquid fill level is lower. This is because when the liquid fill level is lower and the pendulum mass is much smaller, even though the arm length of the pendulum is bigger. It can be deduced that the oscillation amplitude is determined by the liquid sloshing mass to a large degree. This conclusion can also be proved by comparing Figures 9 and 13. Although the liquid fill level of 0.8 is bigger than that of 0.5 shown in Figure 9, the sloshing liquid mass for the former is smaller than that for the latter, as seen from (3) and (4). Therefore, vehicle roll stability when the liquid fill level is 0.5 is worse than that when the liquid fill level is 0.8.

Above all, the tank vehicle's roll stability is greatly influenced when the liquid fill level is near to 0.5, in which case the sloshing liquid mass occupies a large proportion

of the liquid bulk and the mass of the liquid bulk is quite large. In this situation, the tank vehicle's roll stability is much worse than that of normal cars. However, when the liquid fill level is either fairly small or fairly large, the tank vehicle's roll stability is only slightly influenced by transient liquid sloshing. Therefore, a liquid fill level of 0.4–0.6 is the worst load situation for tankers and should be avoided if possible.

### 5. Conclusions

To investigate the driving stability of tank trucks, the paper used an equivalent trammel pendulum for liquid sloshing in a partially filled tank. The trammel pendulum oscillated in the tank and was described in reference to tank-fixed coordinates. To couple the motion of the trammel pendulum with

that of the tank truck, the motion equation of the pendulum under a noninertial coordinate system was analyzed using a Lagrangian function. Based on this, a dynamic model for a tank truck was established using Newton's first law and the angular momentum. A typical tank truck was selected and its driving stability is studied under a steering angle step test. Studying tankers' driving stability and the main influencing factors is of great importance for evaluating tankers' driving safety, as well as for developing active/passive roll control systems for them.

The following important discoveries were found from the driving stability simulations.

- (1) The curves of the tank truck driving variables fluctuate until they return to the steady state. The fluctuation frequency descends with an increase in the liquid fill level while holding the testing situation constant. The fluctuation frequency is close to the pendulum's oscillation frequency. The overshoots of the driving variables depend on the pendulum mass, which is equal to the liquid mass that participates in transient sloshing.
- (2) The pendulum mass occupies a large proportion of the mass of liquid bulk when the liquid fill level is in the range of 0.3–0.7. Also, the mass of liquid bulk is quite large in such cases, the overshooting of the tanker driving variables is quite large, and the tanker's driving stability is greatly impacted by transient liquid sloshing. Therefore, overshoot control for tankers' driving variables must be considered to support their driving safety. Moreover, the situation of a liquid fill level near to 0.5 is the worst state in terms of load and should be avoided.
- (3) The driving stability of a tanker depends on the duration of the driving variables to a large degree.

Since we made the assumption in deriving the inertia tensors for the pendulum mass and the fixed liquid mass that the fixed mass is located at the bottom of the tank and is always static when the tank tilts, the inertia tensor value differs from the real situation, and the simulation results for the tank truck's driving stability will also have been slightly influenced. Thus, a theory and method for physically distinguishing the pendulum's mass and the fixed liquid mass from the liquid bulk will be proposed in a future study.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This research is supported by Research on Evaluation and Detection Technology of Handling and Driving Stability for Commercial Vehicles (no. 2009BAG13A04) and the National Natural Science Foundation of China (Grant nos. 51375200 and 51208225).

### References

- [1] J. Woodrooffe, "Evaluation of dangerous goods vehicle safety performance," Report TP 13678-E, Transport Canada, 2000.
- [2] C. Kwon, "Conditional value-at-risk model for hazardous materials transportation," in *Proceedings of the Winter Simulation Conference (WSC '11)*, pp. 1703–1709, December 2011.
- [3] T. T. Treichel, J. P. Hughes, C. P. L. Barkan et al., "Safety performance of tank cars in accidents: probabilities of lading loss," RSI-AAR Railroad Tank Car Safety Research and Test Project, Association of American Railroads, Washington, DC, USA, 2006.
- [4] M. I. Salem, H. Victor, M. Mucino et al., "Review of parameters affecting stability of partially filled heavy-duty tankers," SAE Technical Paper 1999-01-3709, 1999.
- [5] X. Kang, S. Rakheja, and I. Stiharu, "Cargo load shift and its influence on tank vehicle dynamics under braking and turning," *Heavy Vehicle Systems*, vol. 9, no. 3, pp. 173–203, 2002.
- [6] W. H. Wang, Q. Cao, K. Ikeuchi, and H. Bubb, "Reliability and safety analysis methodology for identification of drivers' erroneous actions," *International Journal of Automotive Technology*, vol. 11, no. 6, pp. 873–881, 2010.
- [7] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [8] F. Solaas and O. M. Faltinsen, "Combined numerical and analytical solution for sloshing in two-dimensional tanks of general shape," *Journal of Ship Research*, vol. 41, no. 2, pp. 118–129, 1997.
- [9] K. Modaresi-Tehrani, S. Rakheja, and I. Stiharu, "Three-dimensional analysis of transient slosh within a partly-filled tank equipped with baffles," *Vehicle System Dynamics*, vol. 45, no. 6, pp. 525–548, 2007.
- [10] L. Strandberg, "Lateral stability of road tankers," VIT Report 138A, National Road and Traffic Research Institute, Linköping, Sweden, 1978.
- [11] R. Ranganathan, S. Rakheja, and S. Sankar, "Steady turning stability of partially filled tank vehicles with arbitrary tank geometry," *Journal of Dynamic Systems, Measurement and Control*, vol. 111, no. 3, pp. 481–489, 1989.
- [12] R. Ranganathan, "Rollover threshold of partially filled tank vehicles with arbitrary tank geometry," *Proceedings of the Institution of Mechanical Engineers D*, vol. 207, no. 3, pp. 241–244, 1993.
- [13] S. Sankar and S. Surial, "Sensitivity analysis approach for fast estimation of rollover stability of heavy articulated vehicles during steady state turning," *Heavy Vehicle Systems*, vol. 1, no. 3, pp. 282–303, 1994.
- [14] C. B. Winkle, "Rollover of heavy commercial vehicles," Report UMTRI-99019, 2001.
- [15] X. Li -S, X. Zheng -L, and H. Liu -F, "Improved algorithm on roll stability evaluation of partially filled tractor-tank semitrailer," *Journal of Jilin University*, vol. 42, no. 5, pp. 1089–1094, 2012 (Chinese).
- [16] K. Modaresi-Tehrani, S. Rakheja, and R. Sedaghati, "Analysis of the overturning moment caused by transient liquid slosh inside a partly filled moving tank," *Proceedings of the Institution of Mechanical Engineers D*, vol. 220, no. 3, pp. 289–301, 2006.
- [17] X. L. Zheng, X. S. Li, and Y. Y. Ren, "Analysis of rollover threshold for partially-filled tank vehicles impacted by transient liquid

- sloshing,” *Journal of Beijing Institute of Technology*, vol. 21, supplement 2, pp. 169–174, 2012.
- [18] F. T. Dodge, “Analytical representation of lateral sloshing by equivalent mechanical models,” in *The Dynamic Behavior of Liquids in Moving Containers*, H. N. Abramson and S. Silverman, Eds., chapter 6, National Aeronautics and Space Administrator, Washington, DC, USA, 1966.
- [19] H. N. Abramson, W. H. Chu, and D. D. Kana, “Some studies of nonlinear lateral sloshing in rigid containers,” *Journal of Applied Mechanics*, vol. 33, pp. 777–784, 1966.
- [20] M. I. Salem, *Rollover Stability of Partially Filled Heavy-Duty Elliptical Tankers Using Trammel Pendulums to Simulate Fluid Sloshing [Ph.D. thesis]*, West Virginia University, Department of Mechanical and Aerospace Engineering, 2000.
- [21] X.-L. Zheng, X.-S. Li, and Y.-Y. Ren, “Equivalent mechanical model for lateral liquid sloshing in partially filled tank vehicles,” *Mathematical Problems in Engineering*, vol. 2012, Article ID 162825, 22 pages, 2012.
- [22] L. Dai, L. Xu, and B. Setiawan, “A new non-linear approach to analysing the dynamic behaviour of tank vehicles subjected to liquid sloshing,” *Proceedings of the Institution of Mechanical Engineers K*, vol. 219, no. 1, pp. 75–86, 2005.

## Research Article

# Dynamic Analysis of Traffic State and Congestion Propagation on Bidirectional Grid Network

Shu-bin Li,<sup>1</sup> Bai-bai Fu,<sup>2</sup> and Jian-feng Zheng<sup>3</sup>

<sup>1</sup> Public Security Department, Shandong Police College, Jinan 250014, China

<sup>2</sup> School of Architecture and Urban Planning, Shandong Jianzhu University, Jinan 250101, China

<sup>3</sup> MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Bai-bai Fu; [fubaibai@163.com](mailto:fubaibai@163.com)

Received 12 July 2013; Revised 22 October 2013; Accepted 29 October 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Shu-bin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Many traffic problems in China such as traffic jams and air pollutions are mainly caused by the increasing traffic volume. In order to alleviate the traffic congestion and improve the network performance, the analysis of traffic state and congestion propagation has attracted a great interest. In this paper, an improved mesoscopic traffic flow model is proposed to capture the speed-density relationship on segments, the length of queue, the flow on links, and so forth. The self-developed dynamic traffic simulation software (DynaCHINA) is used to reproduce the traffic congestion and propagation in a bidirectional grid network for different demand levels. The simulation results show that the proposed model and method are capable of capturing the real traffic states. Hence, our results can provide decision supports for the urban traffic management and planning.

## 1. Introduction

During the past decades, researchers in operations research, transportation engineering, and computer science have taken a keen interest in exploring the problem of traffic congestion and propagation in transportation systems. From the macroview, the congestion formation on a single bottleneck was studied [1–3] and the effect of congestion in a traffic network at the macroscopic system level was reported [4–6]. Particularly, Gentile et al. [5] proposed a new model for the within-day Dynamic Traffic Assignment (DTA) on road networks where the simulation of queue spillovers was explicitly addressed, and the efficiency of the proposed algorithm and its applicability to real instances with large-scale networks was evaluated. At the microscopic system level, there were only a few studies on congestion propagation phenomenon, including the reliability analysis of pedestrian [7] and car-following safety model [8]. Long et al. [9, 10] and Zhang et al. [11] used the cell transmission model to simulate the formation and dissipation of traffic congestion. It was very difficult to analyze the congestion propagation by using the analytical models. This is because the dynamic propagation of traffic

flow in network is extremely complicated, which is related to human behavior (e.g., departure time choice, route choice) and network structure. Therefore, many researchers explore the congestion propagation by making use of the simulation methods. Roberg [12, 13] proposed simulation models and a number of strategies which could be exploited to achieve a controlled dispersion of traffic jams. Roberg and Abbess [14] applied a simulation model to investigate the diagnosis and treatment of traffic jams. Roberg-Orenstein et al. [15] developed several alternative strategies to avoid gridlock and dissipate traffic jams, in terms of the installation of bans at specific network locations. Wright and Roberg [16] proposed a simple analytical model for incident-based jam growth and discussed the effect of the length of the channelized part of roads and stop-line width assignment on jam formation. Wright and Roberg-Orenstein [17] developed simple models for traffic jams and strategies for congestion control on idealized rectangular grid networks.

Most of the mentioned works mainly focus on dynamic of traffic flow, and the congestion propagation characteristics at different demand levels have not been taken into consideration. As we know, several simulation-based DTA systems,

including DynaMIT [18] and DynaSMART-X [19], have been developed for a wide range of Intelligent Transportation System (ITS) applications, and can be used to solve this problem. Unfortunately, these systems cannot be deployed directly in China because of some local traffic characteristics, such as mixed traffic flow situations and trip-maker behavior. In this paper, Dynamic Consistent Hybrid Information based on Network Assignment (DynaCHINA) is used, and the authors are the main member for developing this software.

The paper will be organized as follows. Section 1 introduces the background of the problem and some related studies. Section 2 is the overview of the software DynaCHINA. Section 3 gives dynamic of vehicle movement. Section 4 shows the numerical examples. Section 5 is the conclusion remarks.

## 2. Overview of DynaCHINA

DynaCHINA is a simulation-based real time DTA system that estimates and predicts the current and future traffic conditions. In order to guarantee the credibility of the information to travelers, the guidance provided by DynaCHINA is consistent. Based on the expected response to information, DynaCHINA simulates the traffic conditions that will most likely be experienced by travelers. The historical data, real-time surveillance data, and expected traffic control settings are input data. DynaCHINA can generate prediction-based guidance with respect to departure time, pretrip route and mode choice decisions, and en-route route choice decisions. It supports both prescriptive and descriptive information [18]. Hybrid information means that DynaCHINA optimally uses various data sources (historical, surveillance, O-D data, and floating car data) and hybrid optimization techniques to generate reliable traffic predictions and travel guidance information (such as departure time, mode, and route recommendations) for a variety of information systems and information dissemination strategies in real time. The information provided by DynaCHINA is generated on the basis of dynamic network assignment.

The performance of a simulation-based real-time traffic estimation and prediction system (TrEPS) depends on traffic simulation model which realistically captures vehicular traffic flow dynamics. The origin-destination (OD) flow estimation and prediction methodology incorporate any available information from conventional and emerging traffic data collection technologies (such as automatic vehicle identification (AVI) systems and probe vehicles). In order to address the traffic problems in China, DynaCHINA is different from DynaMIT and DynaSMART, as shown below.

- (1) Anisotropic supply simulator: DynaCHINA uses two main simulation tools, that is, the demand simulator and the supply simulator. The demand simulator is a microscopic simulator that mainly deals with the estimation and prediction of time-dependent OD flows, demand disaggregation to model the socioeconomic characteristics of drivers, and their decisions of departure time and route choice using behavioral models. The supply simulator is a mesoscopic traffic simulator which is used to simulate vehicular movement on the

given network. It can be used to infer traffic flows, queue lengths, speeds, travel times, and densities at all points on the network and thereby serve to indicate network performance [6]. To handle the queue formation and discharge processes, DynaMIT treats moving and queued portions of a link separately, while vehicles in the free-moving portion of the link are moved according to a user-defined macroscopic speed-density relationship. DynaSMART uses virtual queue in conjunction with minimal positive speed. Queue modeling approach in DynaMIT may be conceptually contrived and it is difficult for proper implementation because defining distinct free-moving and queue portions of the traffic stream in general traffic conditions is nontrivial. Transient queues or shockwaves caused by the temporary flow rate transitions may be detected, unless a highly complex set of rules is devised. In DynaSMART, the accuracy of the resultant travel time and link penalty estimation is questionable. Anisotropic mesoscopic model addresses the aforementioned anisotropic property, vehicle moving, and queue formation and discharging in an intuitive and unified manner.

- (2) Availability of floating car data (FCD): both DynaMIT and DynaSMART use the loop detector data to calibrate the model parameters and validate their estimation and prediction capabilities. In the framework for the calibration and validation of DynaCHINA, FCD of 60,000 floating cars are available, in addition to the loop detector data and camera data in Beijing. In case of a positive validation, FCD could make it possible to get the accurate traffic information and travel times on roads where the loops and cameras are missing or out of function. The use of the car itself as a floating traffic sensor is a viable additional information source. FCD is well known as an excellent source of high-quality traffic content. FCD is not a service by itself and it needs to be bundled with other services in an intelligent way (e.g., hybrid navigation). The researchers of DynaCHINA team use the advanced data fusion technologies to integrate and process data from different sources. For example, travel time data from FCD will help calculate route choice probabilities and get more accurate assignment matrix, even set moving speed for some road segments directly instead of using speed-density relationship. DynaCHINA team has the exclusive permission to access FCD in Beijing. Incorporating FCD in the OD estimation problem will improve the effectiveness of the dynamic traffic demand prediction models [19].
- (3) Modeling mixed traffic flow including bicycles, motorcycles, and pedestrians: in China, the mixed-traffic streams in urban networks are composed of motorized, nonmotorized vehicles and pedestrians. These streams comprise standard vehicles, such as cars, buses, and trucks, nonstandard vehicles such as bicycles, motorcycles, and other refitted vehicles as well as pedestrians. For the nonstandard vehicles,

they are slightly restricted by codes, road conditions, and technique criteria. Meanwhile, their traveling objective and routes are even more flexible. Not only there are more differences among the characteristics of each nonstandard vehicle, but also their driver behaviors are obviously diverse from the standard vehicles [20, 21]. At present, there rarely exist models suitable for the analysis of traffic streams with nonstandard vehicles, since the most available models are limited in scope and effectiveness. Inevitably, those existent approaches require improvement. The characteristics and special driver behaviors of nonstandard vehicles in the mixed traffic streams are modeled exogenously, which can help calibrate the model parameters of DynaCHINA, realistically and effectively. Based on Cellular Automation (CA) model and fuzzy logic rules, an approach of simulating vehicle location and a decision-making procedure are adopted. Hence, models which are proper for mixed traffic can be established in DynaCHINA. For the detailed description, please see the related research works of DynaCHINA [22–25].

### 3. Dynamic of Vehicle Movement

The network consists of nodes and links with some loading elements. Each link is divided into segments that capture variations of geometry and traffic conditions along the link. Most segments are defined in advance, and additional segments can be dynamically created to capture the presence of the incidents. Nodes correspond to intersections and the loading elements are ODs.

The complexity of the flows on the network is captured by integrating three classes of models. These classes include capacities associated with roadway features; incidents and intersection controls; deterministic queuing reflecting the effect of bottlenecks; and macroscopic speed-density relationships reflecting uninterrupted flow.

**3.1. Output and Acceptance Capacity.** Each segment has a capacity constraint at the downstream end, referred to as the output capacity. The size of the output capacity depends on the section of the physical characteristics (such as width and slope) or unexpected events and control facilities, and so forth. Each segment is divided into a moving part and a queuing part. The moving part corresponds to the section of the segment where vehicles can move. The queuing part represents vehicles that are queued up. Vehicles are assigned to lanes according to their route to their destination. The determination of output capacities is based on recommendations from the Highway Capacity Manual [25]. When the queue length is equal to the segment length, a spillback occurs on a segment. For the detailed description, please see references [26, 27].

**3.2. Queuing.** Queuing model is a set of models, and each particular queue state (formation, disperse, obstruction, etc.) is described by different models. For example, a vehicle will

join in the queue which is dissipating, and the position at the moment ( $t$ ) is

$$q(t) = q(0) + l(ct - m), \quad (1)$$

where  $q(0)$  is the position of the end of the queue at time  $t = 0$ .  $l$  is the average length of vehicles.  $m$  is the number of moving vehicles between the considered vehicle and the end of the queue at time  $t = 0$ . That is,  $m$  is the number of vehicles that reach the queue before the considered vehicle. Without loss of generality, it is assumed that the position of the upstream end of the segment is 0. The delay of the  $i$ th vehicle in the queue is given by

$$T = \frac{i}{c}, \quad (2)$$

where  $c$  is the output capacity of the segment.

**3.3. Movement.** The speed model is based on the assumption that speed is constant on the upstream section of the segment, followed by a deceleration zone covering a downstream section, where the speed of vehicles varies linearly as a function of the position.

We assume that the position of the upstream end of the segment is 0. Therefore, the downstream end is at position  $L$ , where  $L$  is the length of the segment. The speed function can then be written as

$$v(z) = \begin{cases} v_u, & (0 \leq z \leq L - L_s), \\ \lambda(z - L) + v_d, & (L - L_s < z \leq L), \end{cases} \quad (3)$$

where

$$\lambda = \frac{v_d - v_u}{L_s}. \quad (4)$$

The relationship of the speed density is

$$v_u = \begin{cases} v_f, & \rho \leq \rho_{\max}, \\ v_f \left[ 1 - \left( \frac{\rho - \rho_{\max}}{\rho_{\text{jam}}} \right)^\alpha \right]^\beta, & \rho > \rho_{\max}, \end{cases} \quad (5)$$

where  $v_u$  is the speed at the upstream end of the segment,  $v_a$  is the speed at the downstream end of the segment,  $L_s$  is the length of the deceleration zone,  $v_f$  is the free-flow speed on the segment,  $\rho_{\text{jam}}$  is the jam density,  $\rho_{\max}$  is the maximum permitted density under the free flow speed,  $\rho$  is the density, and  $\alpha, \beta$  are parameters.

**3.4. Vehicles Dynamics.** For the moving parts, vehicle movement is driven by the speed density relationship model, and the vehicles are moved based on the speed [14]. As for the deceleration zone, in the absence of a queue, the vehicle will reach position  $z$  at time  $t(z)$ , given by

$$t(z) = \begin{cases} \frac{1}{\lambda} \log \frac{\lambda_{z+v_u}}{\lambda_{z_0+v_u}}, & \text{if } v_u \neq v_d, \\ \frac{z - z_0}{v_u}, & \text{if } v_u = v_d, \end{cases} \quad (6)$$

where at time  $t = 0$ , a vehicle is at position  $z_0$ . Moreover, the vehicle position at any time  $t$  is given by

$$z(t) = \begin{cases} e^{\lambda t} \left( z_0 + \frac{v_u}{\lambda} \right), & \text{if } v_u \neq v_d, \\ v_u t + z_0, & \text{if } v_u = v_d, \end{cases} \quad (7)$$

where  $\lambda$  is defined by (4).

In the presence of a queue, the position  $z(t)$  of a vehicle at any time  $t$  is given by

$$z(t) = e^{\lambda(t)t} \left( z_0 + \frac{v_u}{\lambda(t)} \right) - \frac{v_u}{\lambda(t)}, \quad (8)$$

where

$$\lambda(t) = \frac{-v_u}{q_0 + l(ct - m)}, \quad (9)$$

where  $q_0$ ,  $l$ ,  $c$ , and  $m$  are defined by (1).

**3.5. Algorithm of Path Generation.** For an urban road network with thousands of connections in DynaCHINA system, it is required to complete the simulation of all vehicles (maybe tens of thousands or hundreds of thousands of vehicles) in a relatively short period of 1–5 minutes, which corresponds to 30–60 minutes in the real road network. A more reasonable approach is to divide the massive complicated issue into two relatively simple issues, such as output problem of an effective path which does not have much requirement on real-time and storage and highly-effective search problem of hyperpath information.

Here is the process of generating the path collection algorithm.

*Step 1.* Set a valid path set  $P = \phi$ .

*Step 2.* Get the set of the effective paths under normal conditions and incorporate in  $P$ :  $P = P \cup k - \text{shortest}(G)$ .

*Step 3.* Let the set of the temporary path  $P_h = P$ . For any path  $p$  in the  $P_h$ , for all  $p \in P_h$ , we implement the following operations.

For any one edge  $l$  in  $p$ , if  $l$  is not operated, we implement the following:

- (a) marked “processed”,
- (b) get the subgraph without the edge  $l$ ,  $G^* = G \setminus l$ ,
- (c) generate a valid path set of  $G^*$  and incorporate it into the  $P$ :  $P = P \cup j - \text{shortest}(G^*)$ .

## 4. Simulation Results

**4.1. Simulation Environment.** In this simulation, we take a square format  $10 \times 10$  network as the traffic network, following the references [9–11]. Two links are connected between two adjacent nodes, and each link has one direction, as shown in Figure 1. It should be pointed out that the ideal transport network is operable, because the mesoscopic traffic flow

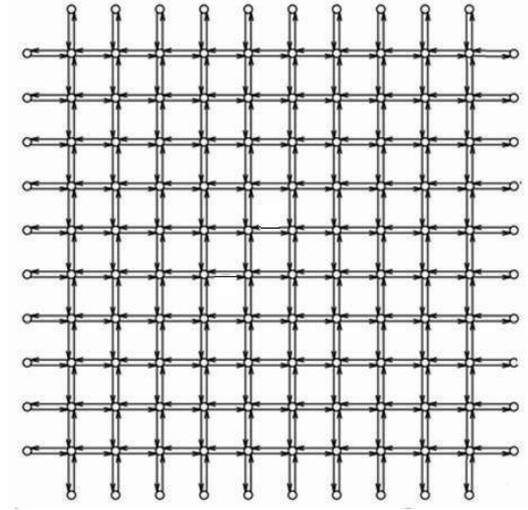


FIGURE 1:  $10 \times 10$  bidirectional grid network.

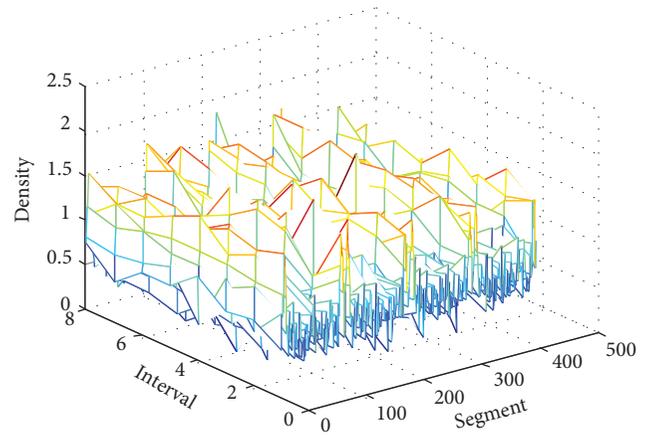


FIGURE 2: Density on each segment per interval when demand is 1.

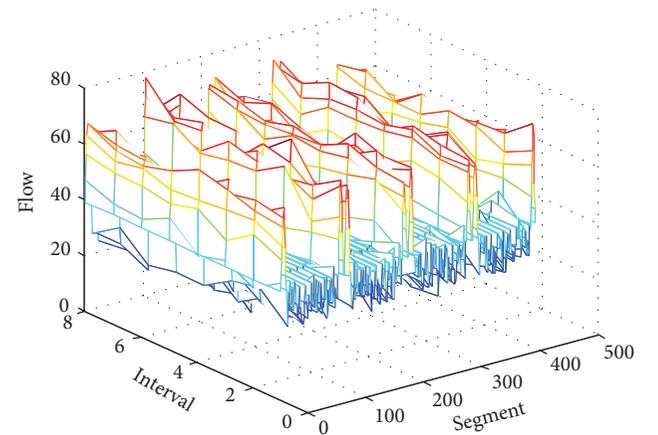


FIGURE 3: Flow on each segment per interval when demand is 1.

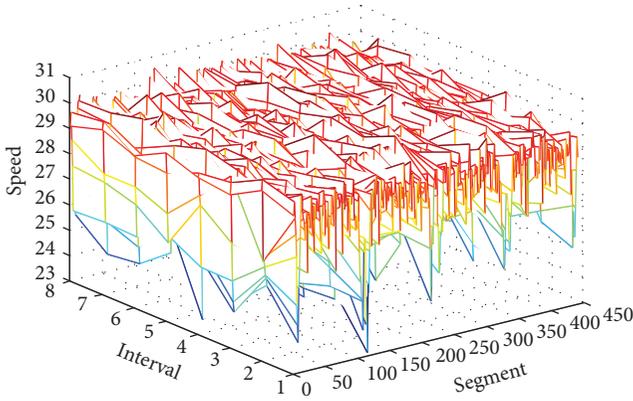


FIGURE 4: Speed on each segment per interval when demand is 1.

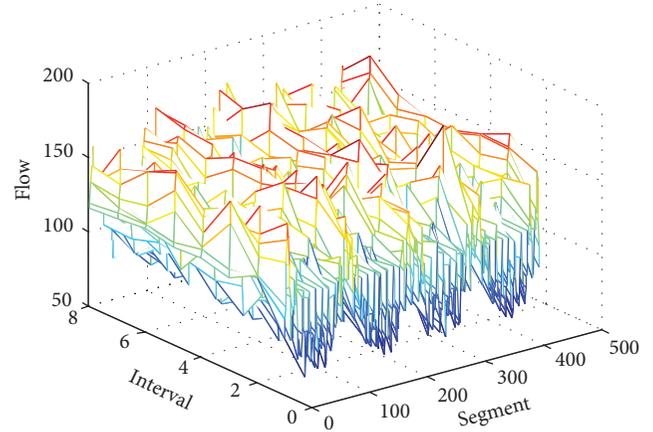


FIGURE 6: Flow on each segment per interval when demand is 3.

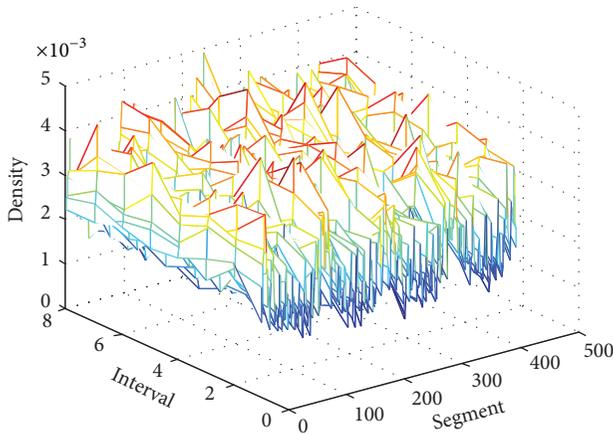


FIGURE 5: Density on each segment per interval when demand is 3.

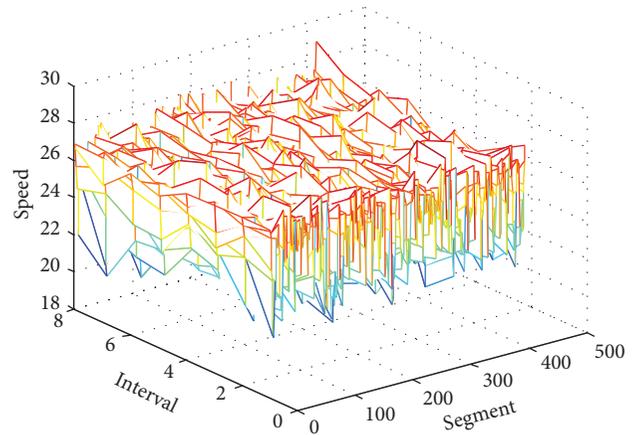


FIGURE 7: Speed on each segment per interval when demand is 3.

model focuses on the vehicle group behavior, and the road network structure has little effect on the content of this study. Each link is defined as 500 meters uniformly, the default is two lanes, and every section is defined as one segment. Each edge node is defined as the transportation zone (i.e., OD points). Jam density is defined as 0.1150 pcu/m/L,  $\alpha$  is 1.9420,  $\beta$  is 0.5040, and the input and output capacity of segment is 0.611 v/s. Without signal control, the designed maximum speed is 60 km/h at all intersections. Simulation time is defined as 15 min, and 8 time periods are considered, leading to a total of two hours of simulation. Network OD demand is loaded with periodic closure boundary condition.

In this paper, the anisotropic mesoscopic simulation modeling concept and model are employed. This is a feature different from other software. The anisotropic mesoscopic simulation modeling concept is defined by the following: for any vehicle  $i$ , its those leading vehicles (in the same lane or in the adjacent lanes) present in vehicle  $i$ 's immediate downstream and within a certain distance are considered to influence vehicle  $i$ 's speed response. By a simple test, at each level of demand, the output average network speed by using DynaCHINA is higher than that using DynaMIT or DynaSMART.

*4.2. Performance for the Network at Different Demand Level.* Generally, the jam density on the segment can be regarded as a benchmark, which is divided into five equally parts, denoted as very smooth, smooth, congestion, very congestion, and jam, respectively. If there are most of links in the network having the same level of jam density, the network can be considered to have the same traffic state.

When the demand of all the OD pairs is set to be 1, the network has not the phenomenon of congestion, which belongs to very smooth. At this demand level, the max vehicle number loaded on the network is 1207. We can see the average density, speed, and flow on each segment at every interval, as shown in Figures 2–4.

When the demand of all the OD pairs is 3, the congestion occurs in some segments, as shown in Figures 6 and 7. The number of vehicles loaded on the network is 2867. Since most segments in this case are not congested, the state of network belongs to smooth.

As can be seen from Figures 2, 3, 4, 5, 6, and 7, the average density and flow are increased very much, but the average speed has some decreased.

When the demand of all the OD pairs is set to be 5, the state of network becomes worse rapidly, which belongs to

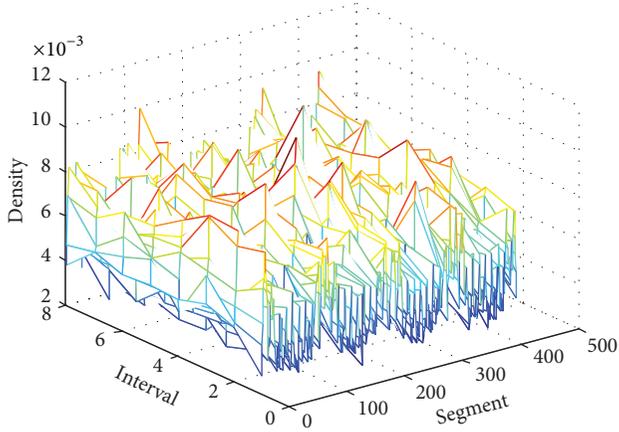


FIGURE 8: Density on each segment per interval when demand is 5.

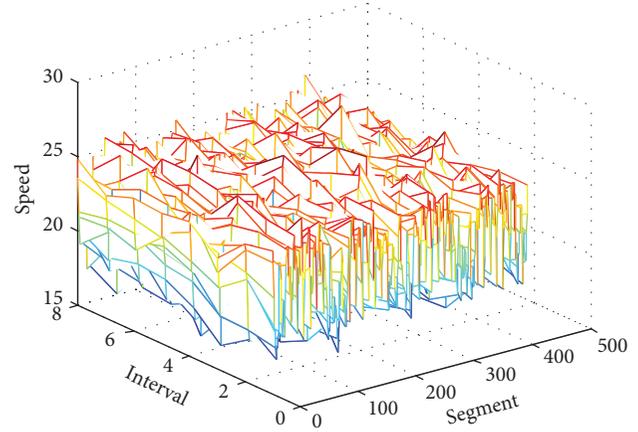


FIGURE 10: Speed on each segment per interval when demand is 5.

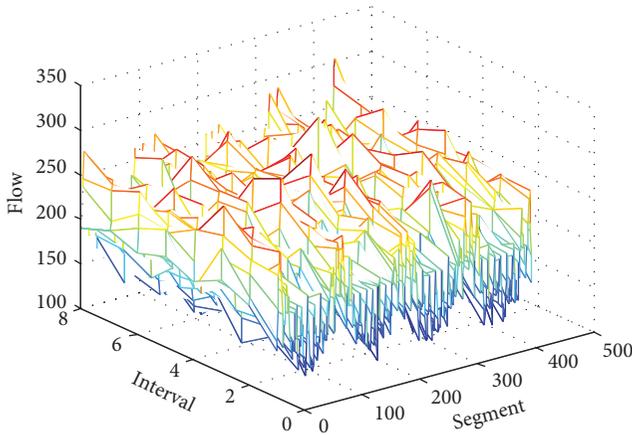


FIGURE 9: Flow on each segment per interval when demand is 5.

very congestion, almost jam state, as shown in Figures 8, 9, and 10. The max number of vehicles loaded on the network is 4386. Most segments are congested in this case. The average density and average flow on the segment become increased sharply, and the average speed is decreased smoothly. If we keep increasing the demand, the network condition becomes worse and worse, leading to the state of extreme congestion. When the demand becomes 1, the loaded vehicle is increased by 780, since we have 40 OD pairs. Another interesting finding is that no matter what the demand is the moving vehicles loaded on the network form a center of symmetry. This is because the OD points are symmetry and the demand is the same in the simulations. However, the kinematic characteristic of vehicles is changeless, and the essential changes for the law of vehicle motion will not occur.

## 5. Concluding Remarks

In this paper, the traffic states and traffic congestion propagation have been studied by developing an improved mesoscopic traffic flow model in the dynamic software of DynaCHINA. The average density, flow, and speed on the segment at every interval for different demand levels are

mainly investigated. Simulation results show that the proposed model and method are capable of capturing the real traffic state. Since the simulations are implemented in a bidirectional grid network which should be improved in the future, the congestion propagation characteristics in the traffic accident should be paid more attention.

## Acknowledgments

This work is partially supported by National Basic Research Program of China (2012CB725400), National Natural Science Foundation of China (71171124, 71131001, and 71222101), Joint Science & Technology of the Ministry of Transport Industry (2009-353-337-480), and Super Computing Technology Special Project of Shandong Province (2011YD01104).

## References

- [1] C. F. Daganzo and J. A. Laval, "Moving bottlenecks: a numerical method that converges in flows," *Transportation Research B*, vol. 39, no. 9, pp. 855–863, 2005.
- [2] G. F. Newell, "A simplified theory of kinematic waves in highway traffic—part III: multi-destination flows," *Transportation Research B*, vol. 27, no. 4, pp. 305–313, 1993.
- [3] D. Ni and J. D. Leonard II, "A simplified kinematic wave model at a merge bottleneck," *Applied Mathematical Modelling*, vol. 29, no. 11, pp. 1054–1072, 2005.
- [4] G. Ziyue and S. Yifan, "A reserve capacity model of optimal signal control with user-equilibrium route choice," *Transportation Research B*, vol. 36, no. 4, pp. 313–323, 2002.
- [5] G. Gentile, L. Meschini, and N. Papola, "Spillback congestion in dynamic traffic assignment: a macroscopic flow model with time-varying bottlenecks," *Transportation Research B*, vol. 41, no. 10, pp. 1114–1138, 2007.
- [6] W. H. K. Lam and Y. Yin, "An activity-based time-dependent traffic assignment model," *Transportation Research B*, vol. 35, no. 6, pp. 549–574, 2001.
- [7] H. Guo, W. Wang, W. Guo, X. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [8] W. H. Wang, W. Zhang, H. W. Guo, H. Bubb, and K. Ikeuchi, "A behavioural car-following safety model for microscopic

- simulation of traffic flow with various driving characteristics,” *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [9] J. C. Long, Z. Y. Gao, H. L. Ren, and A. p. Lian, “Urban traffic congestion propagation and bottleneck identification,” *Science in China F*, vol. 55, no. 7, pp. 948–964, 2008.
- [10] J. C. Long, Z. Y. Gao, P. Orenstein, and H. L. Ren, “Control strategies for dispersing incident-based traffic jams in two-way grid networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 469–481, 2011.
- [11] A. Zhang, Z. Gao, and H. Ren, “Incident-based traffic congestion control strategy,” *Science China Technological Sciences*, vol. 54, no. 5, pp. 1338–1344, 2011.
- [12] P. Roberg, “Development and dispersal of area-wide traffic jams,” *Traffic Engineering & Control*, vol. 35, no. 6, pp. 379–386, 1994.
- [13] P. Roberg, “A distributed strategy for eliminating incident-based traffic jams from urban networks,” *Traffic Engineering & Control*, vol. 36, no. 6, pp. 348–355, 1995.
- [14] P. Roberg and C. R. Abbess, “Diagnosis and treatment of congestion in central urban areas,” *European Journal of Operational Research*, vol. 104, no. 1, pp. 218–230, 1998.
- [15] P. Roberg-Orenstein, C. R. Abbess, and C. Wright, “Traffic jam simulation,” *Journal of Maps*, pp. 107–121, 2007.
- [16] C. Wright and P. Roberg, “The conceptual structure of traffic jams,” *Transport Policy*, vol. 5, no. 1, pp. 23–35, 1998.
- [17] C. Wright and P. Roberg-Orenstein, “Simple models for traffic jams and congestion control,” *Proceedings of the Institution of Civil Engineers*, vol. 135, no. 3, pp. 123–130, 1999.
- [18] M. Ben-Akiva, M. Bierlaire, D. Burton, H. N. Koutsopoulos, and R. Mishalani, “Network state estimation and prediction for real-time transportation management applications,” *Networks and Spatial Economics*, vol. 1, no. 3-4, pp. 293–318, 2001.
- [19] H. S. Mahmassani, “Dynamic network traffic assignment and simulation methodology for advanced systems management applications,” *Networks and Spatial Economics*, vol. 1, no. 3-4, pp. 267–292, 2001.
- [20] C. Antoniou, M. Ben-Akiva, and H. N. Koutsopoulos, “Dynamic traffic demand prediction using conventional and emerging data sources,” *IEE Proceedings*, vol. 153, no. 1, pp. 97–104, 2006.
- [21] T. G. Oketch, “New modeling approach for mixed-traffic streams with nonmotorized vehicles,” *Transportation Research Record*, no. 1705, pp. 61–69, 2000.
- [22] Y. Lin and H. B. Song, “DynaCHINA: specially built real time traffic prediction system for China,” in *Proceedings of the 86th Annual Transportation Research Board Meeting*, Washington, DC, USA, 2007.
- [23] S.-B. Li, Z.-Y. Gao, Y. Lin et al., “Real-time path searching algorithm for large traffic network,” *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 5, pp. 141–147, 2009.
- [24] S. B. Li, J. J. Wu, Z. Y. Gao, Y. Lin, and B. B. Fu, “Bi-dynamic analysis of traffic congestion and propagation based on complex network,” *Acta Physica Sinica*, vol. 60, no. 5, Article ID 050701, 2011.
- [25] Highway Capacity Manual, 1994.
- [26] M. Ben-Akiva, “Development of a deployable real-time dynamic traffic assignment system,” Evaluation Report (Part A): Evaluation of Estimation and Prediction Capabilities, Massachusetts Institute of Technology, 2003.
- [27] M. Ben-Akiva, “Development of a deployable real-time dynamic traffic assignment system,” Evaluation Report (Part B):
- Planning Functionality (DynaMIT-P) and Applications, Massachusetts Institute of Technology, 2003.

## Research Article

# A Hybrid Temporal-Spatio Forecasting Approach for Passenger Flow Status in Chinese High-Speed Railway Transport Hub

Zhengyu Xie,<sup>1,2</sup> Limin Jia,<sup>2</sup> Yong Qin,<sup>2</sup> and Li Wang<sup>1</sup>

<sup>1</sup> School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup> State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Limin Jia; [jialm@vip.sina.com](mailto:jialm@vip.sina.com)

Received 16 July 2013; Revised 4 October 2013; Accepted 18 October 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Zhengyu Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of high-speed railway in China, high-speed railway transport hub (HRTTH) has become the high-density distribution center of passenger flow. In order to accurately detect potential safety hazard hidden in passenger flow, it is necessary to forecast the status of passenger flow. In this paper, we proposed a hybrid temporal-spatio forecasting approach to obtain the passenger flow status in HRTTH. The approach combined temporal forecasting based on radial basis function neural network (RBF NN) and spatio forecasting based on spatial correlation degree. Computational experiments on actual passenger flow status from a specific bottleneck position and its correlation points in HRTTH showed that the proposed approach is effective to forecast the passenger flow status with high precision.

## 1. Introduction

As main influence factors for the safety and sustainability of transportation system, the insecure behaviors and statuses of people are hot issues and difficult problems in traffic safety engineering [1, 2]. With the rapid development of high-speed railway in China, HRTTH has become an interface of multitransportation which includes high-speed railway, civil aviation, highway, waterway, urban rail transit, public transport, and private vehicles, and the safety of passenger flow in HRTTH has attracted more and more attention. As the vital node of passenger transport net, HRTTH is an important collection and distribution center of various transportation modes and massive passenger flow. The distribution quantity of passengers will be sustained to sharply increase with the growth of high-speed railway operation mileage. As the dramatic increase of passengers, high-density passenger flow is generated, which imposed a rigorous challenge to the safety management of HRTTH.

In order to avoid and solve the problems caused by passenger flow abnormal status, many approaches are proposed

in literatures which can be mainly classified into two categories. The first category is the studies on passenger flow modeling and simulating in transport hubs. Gipps and Marksjö [3] focused on the prediction the alteration passenger flow in the passing environment and proposed a model for the interactions between passengers which is intended for use in a graphical computer simulation. Seyfried et al. [4] analyzed the influence of various approaches for the interaction between the passengers on the resulting velocity-density relation based on a modified social force model. Jia et al. [5] analyzed the characteristics of passenger flow and present parameters relation models for passenger flow on different terminal facilities were established based on data statistics. Ji et al. [6] proposed a cell-based model including two steps. The first step is to update speed, which is the cells the passenger can move in one time interval and the other is to analyze the overtaking. Wang et al. [7] simulated the passenger flow in a station hall during the spring festival by modifying the social force model. According to the studies on the first category, decision makers can adjust layout and distribution of facilities and equipment in transport hub to adapt to the modeling and simulating results of passenger

flow. This category studies aim at decreasing and preventing passenger flow abnormal status before they happened.

The other category focuses on the modeling and simulating of congestion evacuations which are caused by high-density passenger flow. Zhong et al. [8] introduced the passenger evacuation design in the construction of metros in China and proposed a dynamic model to simulate the passenger evacuation. Jiang et al. [9] studied the effect of varying maximum upstairs speed and the average minimum width of staircase utilized per person by taking two subway stations in China as examples. VanLandegen and Chen [10] integrated a pedestrian model with a Metrorail transit model to evaluate the performance of the Metrorail in the hypothetical large-scale evacuations. Shi et al. [11] firstly investigated and established the safety strategy of evacuation in metro stations, which involves the occupants needed to be evacuated, the evacuation route, and the safety zone. Based on the studies on the second category, decision makers can make an efficient emergency evacuation decision to determine routes and channels of evacuation. This category studies aim at mitigating passenger flow abnormal status after they happened.

According to the literature review above, most studies focus on solving approaches before or after passenger flow abnormal status happened. Specific literature on real-time changing process of passenger flow status is scarce because of limitation from the difficult acquisition of real-time passenger flow status. With the widespread applications of intelligent video surveillance in Chinese HRTH, real-time acquisition of passenger flow status has become feasible. In this paper, passenger flow status is defined as the amount, velocity, and density of passenger flow. We consider the real-time passenger flow status in bottleneck positions of HRTH and propose a hybrid temporal-spatio forecasting approach to reflect the change of passenger flow status.

The rest of paper is organized as follows: a hybrid temporal-spatio forecasting approach for passenger flow status of bottleneck positions is developed in Section 2. Computational results are reported in Section 3 and finally Section 4 covers the conclusion.

## 2. A Hybrid Temporal-Spatio Forecasting Approach

A hybrid forecasting approach for passenger flow status of bottleneck positions is proposed in this section, which combines temporal and spatio forecasting methods. A temporal forecasting based on RBF NN is proposed to forecast passenger flow status of bottleneck position by using the real-time passenger flow status of the position. The temporal forecasting method can rapidly and precisely reflect the passenger flow status in the bottleneck position but is insensitive for the passenger flow fluctuation from correlation points. So we introduce a spatio forecasting approach based on spatial correlation degree to combine with the temporal forecasting approach for improving the forecasting precision.

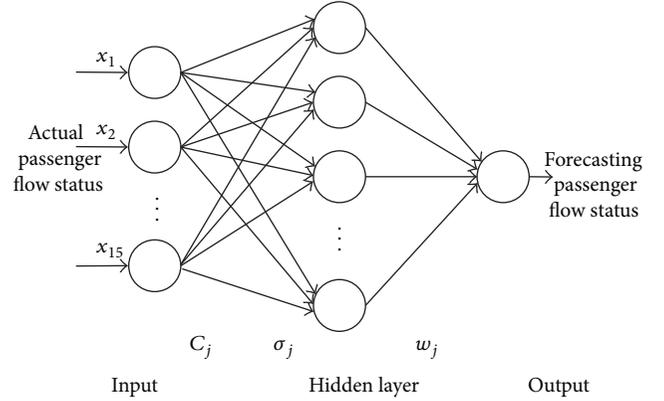


FIGURE 1: Structure of RBF NN.

**2.1. Temporal Forecasting Based on RBF NN.** The RBF NN is a typical feed-forward neural network, which has many merits, such as nonlinear mapping characteristics, self-organized study ability, training fast, and the capability of converging in global optimization and approaching the function in the best way. Simply for its great advantages, RBF NN has been applied in many fields [12–15]. A temporal forecasting approach based on RBF NN is proposed in this section to forecast passenger flow status of bottleneck positions in HRTH.

**2.1.1. Design of RBF NN.** The structure of RBF NN is comprised of three different layers: an input layer, a hidden layer and an output layer. The structure of RBF NN for temporal forecasting is shown in Figure 1.

The input vector  $X = (x_1, x_2, \dots, x_{15})$  is composed of the input signals  $x_i$ , and there are fifteen input neurons in our RBF NN. The output  $h_j(X)$  of the  $j$ th neuron in the hidden neuron in the hidden layer is described by

$$h_j(X) = \varphi(\|X - C_j\|), \quad (1)$$

where  $\varphi(\cdot)$  is the radial basis function,  $C_j(C_j \in R^n)$  is the center of the  $j$ th hidden neuron, and  $\|\cdot\|$  is the Euclidean norm. In our RBF NN,  $\varphi(\cdot)$  is adopted to be the Gaussian function, and  $h_j(X)$  is described by

$$h_j(X) = \exp\left(-\frac{\|X - C_j\|^2}{2\sigma_j^2}\right), \quad (2)$$

where  $\sigma_j$  is the spread of the  $j$ th hidden neuron for the  $i$ th input signal.

The output  $Y(X)$ , which is designed as passenger flow status forecasting, is the linearly combined signal of the outputs  $h_j(X)$  from the hidden layer with the synaptic weights  $w_j$  described by

$$Y(X) = \sum_{j=1}^m w_j h_j(X), \quad (3)$$

where  $m$  is the number of hidden neurons.

TABLE 1: Mapping structure between input variables and output variable.

Input variables	Output variable
$x_1, x_2, \dots, x_{15}$	$x_{16}$
$x_2, x_3, \dots, x_{16}$	$x_{17}$
$\vdots$	$\vdots$
$x_{l-15}, x_{l-14}, \dots, x_{l-1}$	$x_l$

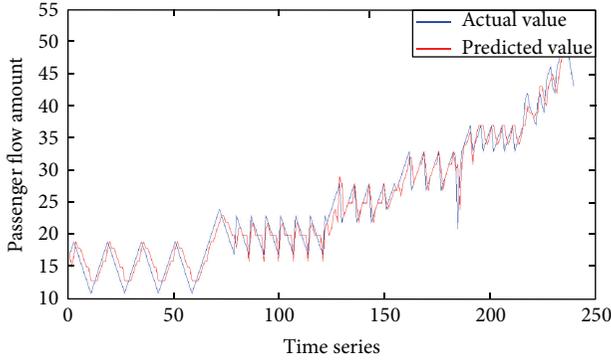


FIGURE 2: Comparison between actual value and forecasting value based on RBF NN.

2.1.2. *Input Variables Determined.* It is very important to select proper input variables for a neural network. On the one hand, as the number of the input variables increases, the NN architecture will be larger and the computing time will be longer. On the other hand, the irrelevant or mutually correlated input variables are not useful for improving the prediction accuracy. Therefore, how to select a few but sufficient input variables is a key issue.

According to plenty of experiments on the input variables, we choose passenger flow status (amount, velocity, or density) of 15 points in the time series as the variables input our RBF NN and passenger flow status (amount, velocity, or density) of the point after the 15th point as the output variable.  $X = \{x_i \mid x_i \in N^+, i = 1, 2, \dots, l\}$  is a time series set and  $x_i$  is the passenger flow status (amount, velocity, or density) of the  $i$ th point in the time series. The mapping structure between input variables and output variable is shown in Table 1.

2.1.3. *RBF NN Testing.* In order to test the accuracy of temporal forecasting, we choose 250 actual passenger flow amounts of one bottleneck position in time series as testing samples. The input and output values after normalization are shown in Table 2.

A comparison between actual value and forecasting value based on RBF NN is shown in Figure 2. The average precision of forecasting is 94.14%. The testing result indicates that the RBF NN has a desirable performance on forecasting.

2.2. *Spatio Forecasting Based on Spatial Correlation Degree.* According to the spatial correlation degree and passenger flow status of bottleneck position  $P_0$  and its correlation points

TABLE 2: Input and output values.

No.	Input value				Output value
1	-0.7561	-0.8537	0.9024	0.8537	-0.7561
	-1.0000	-0.9512	-0.9024	-0.8537	
	-0.8049	-0.7561	-0.7073	-0.6585	
	-0.6098	-0.6585	-0.7073		
2	-0.8537	0.9024	0.8537	-1.0000	-0.8049
	-0.9512	-0.9024	-0.8537	-0.8049	
	-0.7561	-0.7073	-0.6585	-0.6098	
	-0.6585	-0.7073	-0.7561		
3	0.9024	0.8537	-1.0000	-0.9512	-0.8537
	-0.9024	-0.8537	-0.8049	-0.7561	
	-0.7073	-0.6585	-0.6098	-0.6585	
	-0.7073	-0.7561	-0.8049		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
234	0.5122	0.4146	0.3659	0.6098	0.6585
	0.6585	0.7073	0.5610	0.5122	
	0.7073	0.7561	0.9512	0.9024	
	1.0000	0.8537	0.7561		
235	0.4146	0.3659	0.6098	0.6585	0.5610
	0.7073	0.5610	0.5122	0.7073	
	0.7561	0.9512	0.9024	1.0000	
	0.8537	0.7561	0.6585		

$C_i$ , a spatio forecasting model is proposed to forecast the passenger flow status of bottleneck position  $P_0$ .

2.2.1. *Analysis of Spatial Correlation Degree.*  $X_i = \{x_i(k) \mid k = 1, 2, \dots, n\}$  ( $i = 1, 2, \dots, m$ ) is a sequence set,  $x_0(k)$  is a reference sequence, and  $x_i(k)$  is comparison sequence.  $r(x_0(k), x_i(k))$  is the correlation coefficient of comparison sequence relative to reference sequence at period  $k$ , as shown in (4). Consider

$$r(x_0(k), x_i(k)) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \zeta \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \zeta \max_i \max_k |x_0(k) - x_i(k)|}, \quad (4)$$

where  $\zeta$  is the discriminating coefficient,  $\zeta \in (0, 1)$ .

$r(x_0, x_i)$  is the correlation degree of the  $i$ th comparison sequence  $x_i$  relative to reference sequence  $x_0$ , as shown in (5). Consider

$$r(x_0, x_i) = \frac{1}{n} \sum_{k=1}^n r(x_0(k), x_i(k)). \quad (5)$$

$\alpha_i$  is the correlation weight coefficient and is described by (6). Consider

$$\alpha_i = \frac{r(x_0, x_i)}{\sum_{j=1}^n r(x_0, x_j)}. \quad (6)$$

2.2.2. *Spatio Forecasting Model.* According to the passenger flow status of  $C_i$  and the spatio correlation degree between  $P_0$

and  $C_i$ , the passenger flow status of  $P_0$  is forecasted based on spatio forecasting model. The initial forecasting value  $x_c'(t)$  of  $P_0$  at period  $t$  is calculated by (7). Consider

$$x_c'(t) = \sum_{i=1}^n \alpha_i x_i(t - t_i), \quad (7)$$

where  $x_i(t)$  is passenger flow status of  $C_i$  at period  $t$ ,  $t_i$  is the average moving time from  $C_i$  to  $P_0$ , and  $\alpha_i$  is the correlation weight coefficient between correlated points  $C_i$  and congested point  $P_0$ . According to (6),  $\alpha_i$  is described by

$$\alpha_i = \frac{r(P_0, C_i)}{\sum_{j=1}^n r(P_0, C_j)}, \quad (8)$$

where  $r(P_0, C_i)$  is the correlation degree of between  $C_i$  and  $P_0$ .

In order to reduce the errors which are caused by uncertainties passenger flow status, we use the change of initial forecasting value  $x_c'(t)$  and actual passenger flow status value  $x_c(0)$  of  $P_0$  at the beginning of period  $t$  to describe the forecasting value  $x_c(t)$  of  $P_0$  at period  $t$ . The spatio forecasting model of passenger flow status is shown in (9). Consider

$$x_c(t) = \sum_{i=1}^n \alpha_i x_i(t - t_i) - \sum_{i=1}^n \alpha_i x_i(0) + x_c(0). \quad (9)$$

The velocity in correlated points prominently affects the density of congested point; the faster the velocity of correlated points, the weaker affection on the density of congested point. So we adopt velocity to improve the density forecasting. The spatio forecasting model of passenger flow density is shown in (10). Consider

$$x_{\rho c}(t) = \sum_{i=1}^n \frac{k\nu_i(t - t_i)}{\nu_0(t - t_i)} \alpha_i x_{\rho i}(t - t_i) - \sum_{i=1}^n \alpha_i x_{\rho i}(0) + x_{\rho}(0), \quad (10)$$

where  $x_{\rho c}(t)$  is the forecasting passenger flow density of  $P_0$  at period  $t$ ,  $\nu_i(t - t_i)$  is the velocity of correlated points at period  $t - t_i$ ,  $\nu_0(t - t_i)$  is the velocity of congested point at period  $t - t_i$ , and  $x_{\rho}(0)$  is the actual density of  $P_0$  at the beginning of period  $t$ .

Similarly, the density in correlated points prominently affects the velocity of congested point; the higher the density of correlated points, the faster the velocity of congested point. So, we adopt density to improve the velocity forecasting. The spatio forecasting model of passenger flow velocity is shown in (11). Consider

$$x_{\nu c}(t) = \sum_{i=1}^n \frac{k\rho_0(t - t_i)}{\rho_i(t - t_i)} \alpha_i x_{\nu i}(t - t_i) - \sum_{i=1}^n \alpha_i x_{\nu i}(0) + x_{\nu}(0), \quad (11)$$

where  $x_{\nu c}(t)$  is the forecasting passenger flow velocity of  $P_0$  at period  $t$ ,  $\rho_i(t - t_i)$  is the density of correlated points at period  $t - t_i$ ,  $\rho_0(t - t_i)$  is the density of congested point at period  $t - t_i$ , and  $x_{\nu}(0)$  is the actual velocity of  $P_0$  at the beginning of period  $t$ .

**2.3. Temporal-Spatio Forecasting Combination.** A hybrid forecasting approach to combine the spatio and temporal forecasting is proposed in this section. The forecasting value  $x_r(t)$  based on RBF NN and the forecasting value  $x_c(t)$  based on spatioforecasting model are combined by a linear regression model, which is shown in (12). Consider

$$f(t) = \beta_1 x_r(t) + \beta_2 x_c(t) + \beta_3. \quad (12)$$

The parameters of the model are determined by numerical fitting of actual passenger flow status values and forecasting values generated by spatio and temporal forecasting methods.

According to the spatio and temporal forecasting methods and the combination model, the temporal-spatio forecasting model of passenger flow density is described by

$$f_{\rho}(t) = \beta_{\rho 1} x_{\rho r}(t) + \beta_{\rho 2} \left[ \sum_{i=1}^n \frac{k\nu_i(t - t_i)}{\nu_0(t - t_i)} \alpha_i x_{\rho i}(t - t_i) - \sum_{i=1}^n \alpha_i x_{\rho i}(0) + x_{\rho}(0) \right] + \beta_{\rho 3}. \quad (13)$$

According to the spatio and temporal forecasting methods and the combination model, the temporal-spatio forecasting model of passenger flow velocity is described by

$$f_{\nu}(t) = \beta_{\nu 1} x_{\nu r}(t) + \beta_{\nu 2} \left[ \sum_{i=1}^n \frac{k\rho_0(t - t_i)}{\rho_i(t - t_i)} \alpha_i x_{\nu i}(t - t_i) - \sum_{i=1}^n \alpha_i x_{\nu i}(0) + x_{\nu}(0) \right] + \beta_{\nu 3}. \quad (14)$$

### 3. Computational Experiments

To illustrate the proposed forecasting approach, computational experiments are performed by using the actual passenger flow status from a specific bottleneck position in the Chinese HRTM. In order to assess the improvement of our approach, the forecasting of passenger flow density and velocity among temporal forecasting based on RBF NN, spatio forecasting based on spatial correlation degree, and hybrid temporal-spatio forecasting approach are compared. These numerical experiments are performed based on a personal computer with Intel Core(TM) i5-2450M @ 2.50 GHz processors and 4 GB RAM.

To implement the proposed forecasting approach, the parameters related to specific bottleneck position of HRTM are needed. The specific bottleneck position is a ticket entrance  $P$  with 3 correlation points  $A$ ,  $B$ , and  $C$ . The average passenger moving time of  $A$  to  $P$  is 35 s,  $B$  to  $P$  is 55 s, and  $C$  to  $P$  is 20 s. We choose 311 data of passenger flow density and 311 data of passenger flow velocity as the computational examples.

According to RBF NN designed in Section 2, temporal forecasting experiments on passenger flow density and

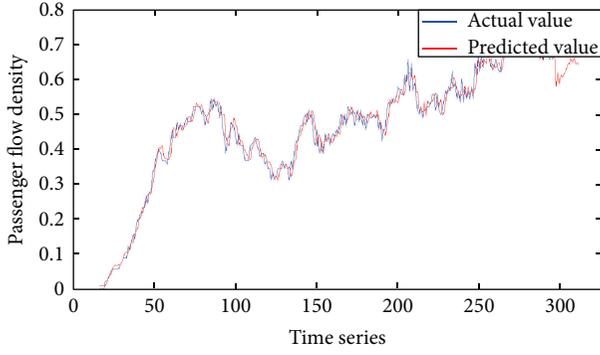


FIGURE 3: Comparison between actual value and forecasting value of passenger flow density based on RBF NN.

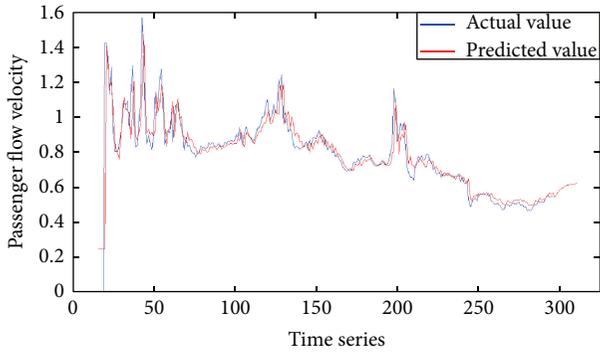


FIGURE 4: Comparison between actual value and forecasting value of passenger flow velocity based on RBF NN.

velocity are implemented. The comparison between actual value and forecasting value of passenger flow density based on RBF NN is shown in Figure 3 and comparison between actual value and forecasting value of passenger flow velocity based on RBF NN is shown in Figure 4.

According to spatio forecasting model proposed in Section 2, spatio forecasting experiments on passenger flow density and velocity are implemented.  $x_0$  is the passenger flow status sequence of  $P$ ,  $x_1$  is the passenger flow status sequence of  $A$ ,  $x_2$  is the passenger flow status sequence of  $B$ , and  $x_3$  is the passenger flow status sequence of  $C$ . The spatial correlation degrees are calculated according to (4) and (5). The calculation result is shown as follows:

$$r(0,1) = 0.510, \quad r(0,2) = 0.761, \quad r(0,3) = 0.404. \quad (15)$$

The forecasting value of passenger flow density based on spatio forecasting model is calculated by (16) and the

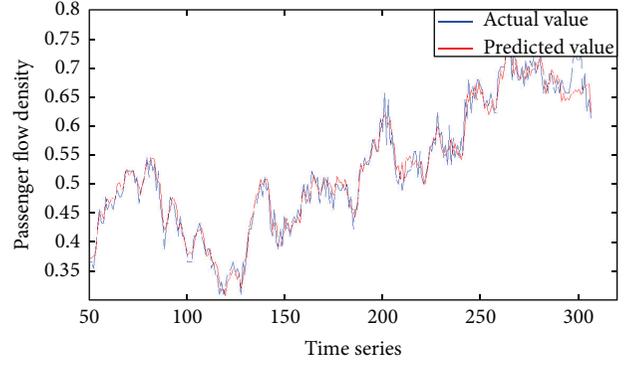


FIGURE 5: Comparison between actual value and forecasting value of passenger flow density based on spatio forecasting model.

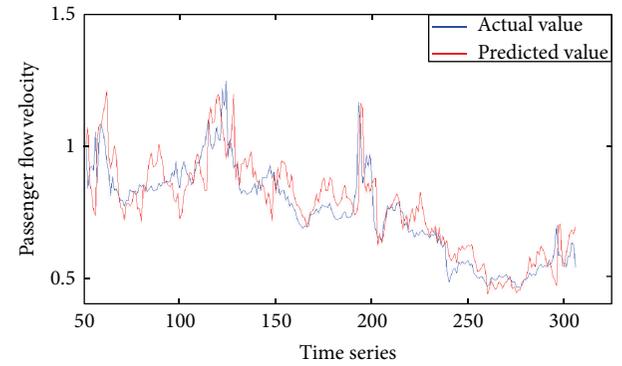


FIGURE 6: Comparison between actual value and forecasting value of passenger flow velocity based on spatio forecasting model.

forecasting value of passenger flow velocity based on spatio forecasting model is calculated by (17). Consider

$$\begin{aligned} x_{\rho c}(t) &= \sum_{i=1}^n \frac{k\nu_i}{\nu_0} \alpha_i x_{\rho i}(t-t_i) - \sum_{i=1}^n \alpha_i x_{\rho i}(0) + x_{\rho}(0) \\ &= \frac{0.304\nu_1}{\nu_0} x_{\rho 1}(t-35) + \frac{0.454\nu_2}{\nu_0} x_{\rho 2}(t-55) \\ &\quad + \frac{0.241\nu_3}{\nu_0} x_{\rho 3}(t-20) + 0.206, \end{aligned} \quad (16)$$

$$\begin{aligned} x_{\nu c}(t) &= \sum_{i=1}^n \frac{k\rho_0}{\rho_i} \alpha_i x_{\nu i}(t-t_i) - \sum_{i=1}^n \alpha_i x_{\nu i}(0) + x_{\nu}(0) \\ &= \frac{0.213\rho_1}{\rho_0} x_{\nu 1}(t-35) + \frac{0.363\rho_2}{\rho_0} x_{\nu 2}(t-55) \\ &\quad + \frac{0.193\rho_3}{\rho_0} x_{\nu 3}(t-20) + 0.671. \end{aligned} \quad (17)$$

The comparison between actual value and forecasting value of passenger flow density based on spatio forecasting model is shown in Figure 5 and comparison between actual value and forecasting value of passenger flow velocity based on spatio forecasting model is shown in Figure 6.

TABLE 3: Average forecasting precision comparison of three approaches.

Forecasting approach	Average forecasting precision of density	Average CPU time of density forecasting (s)	Average forecasting precision of velocity	Average CPU time of velocity forecasting (s)
Temporal forecasting approach based on RBF NN	95.21%	9.8	94.73%	9.6
Spatio forecasting approach based on spatial correlation degree	91.23%	11.2	88.21%	12.1
Hybrid temporal-spatio forecasting approach	96.83%	12.9	96.10%	13.7

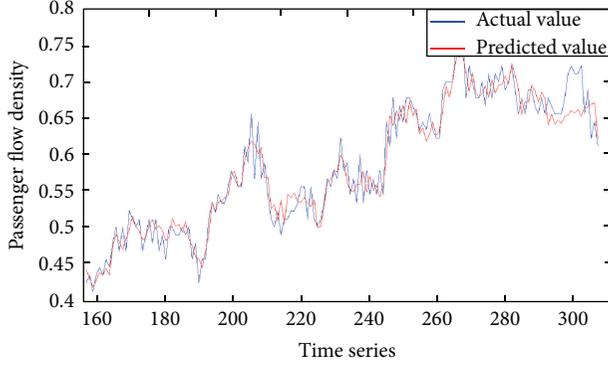


FIGURE 7: Comparison between actual value and forecasting value of passenger flow density based on hybrid temporal-spatio forecasting approach.

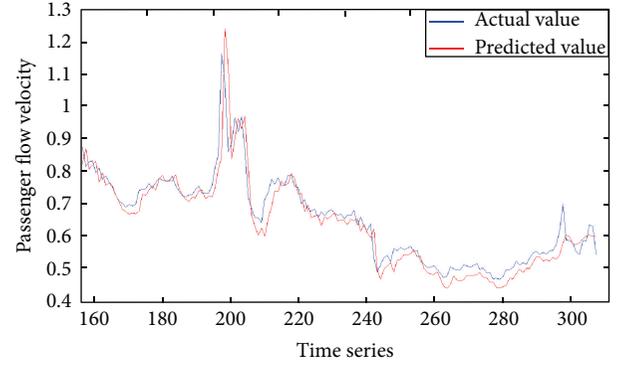


FIGURE 8: Comparison between actual value and forecasting value of passenger flow velocity based on hybrid temporal-spatio forecasting approach.

Based on the numerical fitting of actual and forecasting passenger flow status value in  $P$ , the forecasting value of passenger flow density based on hybrid temporal-spatio forecasting approach is calculated by (18) and the forecasting value of passenger flow velocity based on hybrid temporal-spatio forecasting approach is calculated by (19). Consider

$$x_{\rho}(t) = 0.889x_{\rho r}(t) + \frac{0.034\nu_1}{\nu_0}x_{\rho 1}(t-25) + \frac{0.050\nu_2}{\nu_0}x_{\rho 2}(t-55) \quad (18)$$

$$+ \frac{0.0271\nu_3}{\nu_0}x_{\rho 3}(t-20) + 0.022,$$

$$x_{\nu}(t) = 0.903x_{\nu r}(t) + \frac{0.020\rho_1}{\rho_0}x_{\nu 1}(t-35) + \frac{0.035\rho_2}{\rho_0}x_{\nu 2}(t-55) + \frac{0.019\rho_3}{\rho_0}x_{\nu 3}(t-20) + 0.044. \quad (19)$$

The comparison between actual value and forecasting value of passenger flow density based on hybrid temporal-spatio forecasting approach is shown in Figure 7 and comparison between actual value and forecasting value of passenger flow velocity based on hybrid temporal-spatio forecasting approach is shown in Figure 8.

The average forecasting precision comparison of three approaches mentioned above is shown in Table 3.

As observed in Table 3, the hybrid temporal-spatio forecasting approach proposed in this paper has the higher forecasting precisions than temporal and spatio forecasting approaches under the similar computation complexity. The results of computational experiments indicate that our approach is effective to forecast the passenger flow status of bottleneck positions in Chinese HRTH.

#### 4. Conclusion

In this paper, we considered the forecasting approach for passenger flow status in the Chinese HRTH. A hybrid temporal-spatio forecasting approach was proposed, which combined temporal forecasting and spatio forecasting. The temporal forecasting based on RBF NN could fast and accurately forecast the status change of passenger flow but was insensitive for the influences from correlation points. A spatio forecasting approach based on spatial correlation degree was introduced to combine with the temporal forecasting approach to avoid the influences and improve the forecasting precision. Computational experiments on the actual passenger flow density and velocity from a specific bottleneck position and its correlation points in Chinese HRTH showed that the approach proposed in this paper is effective to forecast the passenger flow status of bottleneck position in HRTH with high forecasting precision for different types of passenger flow status. In the future, considering the passenger flow abnormal status forewarning of bottleneck position

based on the passenger flow status forecasting is a possibility for further research.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgment

This research was supported by a Grant (no. I11A300010) from the National Natural Science Foundation of China.

### References

- [1] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [2] H. Guo, W. Wang, W. Guo, X. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [3] P. G. Gipps and B. Marksjö, "A micro-simulation model for pedestrian flows," *Mathematics and Computers in Simulation*, vol. 27, no. 2-3, pp. 95–105, 1985.
- [4] A. Seyfried, B. Steffen, and T. Lippert, "Basics of modelling the pedestrian flow," *Physica A*, vol. 368, no. 1, pp. 232–238, 2006.
- [5] H.-F. Jia, L.-L. Yang, and M. Tang, "Pedestrian flow characteristics analysis and model parameter calibration in comprehensive transport terminal," *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 5, pp. 117–123, 2009.
- [6] X. Ji, X. Zhou, and B. Ran, "A cell-based study on pedestrian acceleration and overtaking in a transfer station corridor," *Physica A*, vol. 392, no. 8, pp. 1828–1839, 2013.
- [7] L. Wang, Q. Zhang, Y. Cai, J. L. Zhang, and Q. G. Ma, "Simulation study of pedestrian flow in a station hall during the spring festival travel rush," *Physica A*, vol. 392, no. 10, pp. 2470–2478, 2013.
- [8] M. Zhong, C. Shi, X. Tu, T. Fu, and L. He, "Study of the human evacuation simulation of metro fire safety analysis in China," *Journal of Loss Prevention in the Process Industries*, vol. 21, no. 3, pp. 287–298, 2008.
- [9] C. S. Jiang, F. Yuan, and W. K. Chow, "Effect of varying two key parameters in simulating evacuation for subway stations in China," *Safety Science*, vol. 48, no. 4, pp. 445–451, 2010.
- [10] L. D. VanLandegen and X. Chen, "Microsimulation of large-scale evacuations utilizing metrorail transit," *Applied Geography*, vol. 32, no. 2, pp. 787–797, 2012.
- [11] C. Shi, M. Zhong, X. Nong, L. He, J. Shi, and G. Feng, "Modeling and safety strategy of passenger evacuation in a metro station in China," *Safety Science*, vol. 50, no. 5, pp. 1319–1332, 2012.
- [12] J. Bi, S. Shao, W. Guan, and L. Wang, "State of charge estimation of Li-ion batteries in an electric vehicle based on a radial-basis-function neural network," *Chinese Physics B*, vol. 21, no. 11, Article ID 118801, 2012.
- [13] W. Y. Chang, "Equivalent circuit parameters estimation for PEM fuel cell using RBF neural network and enhanced particle swarm optimization," *Mathematical Problems in Engineering*, vol. 2013, Article ID 672681, 8 pages, 2013.
- [14] E. E. Abusham and E. K. Wong, "Locally linear discriminate embedding for face recognition," *Discrete Dynamics in Nature and Society*, vol. 2009, Article ID 916382, 8 pages, 2009.
- [15] J.-D. Wu and J.-C. Liu, "A forecasting system for car fuel consumption using a radial basis function neural network," *Expert Systems with Applications*, vol. 39, no. 2, pp. 1883–1888, 2012.

## Research Article

# Exploring the Effects of Different Walking Strategies on Bi-Directional Pedestrian Flow

Lili Lu,<sup>1,2</sup> Gang Ren,<sup>1,2</sup> Wei Wang,<sup>1,2</sup> Chen Yu,<sup>1,2</sup> and Chenzi Ding<sup>1,2</sup>

<sup>1</sup> Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou No. 2, Nanjing 210096, China

<sup>2</sup> Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Si Pai Lou No. 2, Nanjing 210096, China

Correspondence should be addressed to Gang Ren; [rengang@seu.edu.cn](mailto:rengang@seu.edu.cn)

Received 21 August 2013; Accepted 8 October 2013

Academic Editor: Niu Huimin

Copyright © 2013 Lili Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Three types of different walking behaviors (right preference, conformity, and space priority) are taken into account to model bi-directional pedestrian flow in the channel with cellular-automata formulation. The fundamental diagrams of *R*-pedestrian flow, *C*-pedestrian flow, and *S*-pedestrian flow are obtained from the simulation result to analyze the effect of these behaviors on bi-direction flow. The *C*-pedestrian flow has the minimum critical density and *R*-pedestrian flow has the highest, while the *S*-pedestrian flow has higher average-speed than other two types of pedestrian flow under the same density. Further, through the study of pedestrian distribution in the channel and the proportion of pedestrians not able to move to the front cell, reasons leading to different characteristics of these three types of pedestrian flow are analyzed. Moreover, the simulation experiment based on BehaviorSearch is designed to explore the optimal percentages of *R*-pedestrian, *C*-pedestrian, and *S*-pedestrian in pedestrian flow. The result of the experiment shows that the condition that makes the highest average speed of pedestrian flow is not that pedestrian flow consists of purely one type of pedestrians, but pedestrian flow mixed with *S*-pedestrians as majority and *C*-pedestrians and *R*-pedestrians as minority.

## 1. Introduction

In recent years, pedestrian flow research has found great interest in many research fields [1–3]. The complex behaviors of pedestrian can emerge a variety of interesting self-organization phenomena, such as the oscillations at bottlenecks [4], the lane formation in pedestrian bi-directional flow, or the turbulent movement in dense crowds [5]. It is known to us that understanding the behavior of pedestrian is fundamental to developing a predictable model for the design of urban infrastructures, traffic management, or crowd safety during mass events or evacuation processes. However, pedestrians' behavior movement is flexible and changeable according to walking conventions, psychology, environment, and so forth. Thus, it is a great challenge to model pedestrian behavior accurately.

Recently, many models of pedestrian behavior have been developed to simulate pedestrian flow including social-force model [6–8], lattice gas model [9], cellular-automata (CA) model [10–13], and agent-based models [14]. Particularly,

cellular automation has been widely used to simulate pedestrian movement, in which an approximate actual pedestrian behavior can be described by setting simple local rules to each individual, and the collective behavior of pedestrians emerges as an outgrowth of microsimulation rule set. The collective phenomena of pedestrian flow like the best known spontaneous formation of lanes in bi-directional pedestrian flow and oscillatory changes of the walking direction at narrow passages have been successfully simulated by CA approach due to the computational efficiency. To achieve more realistic simulations, various CA models begin to consider pedestrians' sociological and psychological characteristic. Researchers are paying more attention to pedestrians' behavior to enhance the model and to study the effects of different behavior on pedestrian flow operation.

Among the existing bi-directional pedestrian CA models, the following sidling and walking back behavior were considered and their effects on pedestrian dynamic were discussed such that the back-step strategy can reduce the jamming condition significantly [15–17]. On the other hand, some

researchers focused on the individual movement behavior influenced by the interaction or relationships between other individuals. Ma et al. (2010) built the KNN CA model based on the  $k$ -nearest-neighbor interaction pattern that pedestrian's direction of choosing behavior was effected by the distribution of fixed neighbors, which presented self-organization phenomena and validated that the  $k$ -nearest neighbors play a fundamental role in the emergency of pedestrian collective behavior [18]. Wang et al. (2012) proposed a CA model to simulate team moving behavior and conducted the simulation experiment to discuss the effect of this type of behavior. It was found that the capacity of channel would decrease if the teeming number increased [19].

However, the models mentioned above mainly concentrated on certain type of pedestrian behavior and ignored the complexity and randomness of pedestrian behavior. Thus, in this paper we present a bi-directional pedestrian model where each individual may take different walking strategies while walking. This modified model can be used to analyze the effect of different moving strategies on pedestrian flow through the simulation method, which will be helpful to understand the operating mechanism of pedestrian flow and to put forward measures to improve its operating efficiency.

What is more, for simulation experiment, a challenging task is to vary the parameters of the model continuously and discover the impact of different parameter setting on simulation results. For example, if we want to check whether the critical density is independent of the system size as many pedestrian simulations have done, we need to run the model with all combinations of parameters setting exhaustively. It is generally time consuming and not feasible. So in this paper, the new soft tool BehaviorSearch is introduced and applied to explore the parameter space of model, and thus a new method and idea of performing this work for pedestrian flow simulation is provided [20].

The major objective of this paper is to study the effects of different walking strategies on bi-directional pedestrian flow. The remainder of this paper is organized as follows. The next section presents the process of modeling pedestrians with three different walking strategies using CA approach; Section 3 provides the simulation results to analyze the effect of these walking strategies on bi-directional pedestrian flow. At the same time, the simulation experiment based on BehaviorSearch for the purpose of exploring parameter space is described, which will help us gain insight into the relationship between pedestrian behavior and pedestrian flow operation. Finally, the major findings and conclusions are summarized and the direction for future work is provided in the last section.

## 2. Model Development

**2.1. Initialization.** In this model, pedestrian moving space is portioned into  $W \times L$  grids in the plane. Every grid is a cell with the size of  $0.4 \text{ m} \times 0.4 \text{ m}$  [18], which is the typical space occupied by a pedestrian in a dense crowd. Each cell must be occupied by one pedestrian or must be empty. There are two types of pedestrian in the system. One is moving to the right, and the other is moving to the left, as is shown in Figure 1.

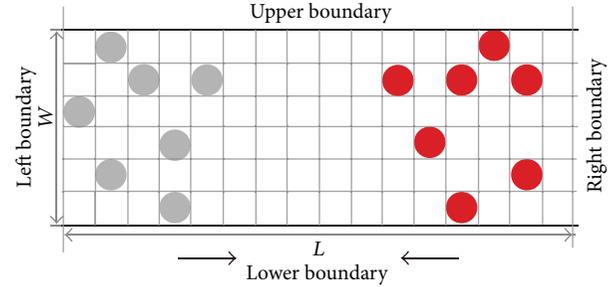


FIGURE 1: Schematic illustration of the system of pedestrian counterflow.

Every pedestrian can move to an empty grid or stay still at each time step. At the initial time step, all pedestrians are randomly distributed on the  $W \times L$  grids with a given density.

The upper and lower boundaries of the system are closed, while the left and right ones are periodical, meaning that when the right-moving pedestrians reach the right boundary, they will move back from the left boundary, and when the left-moving pedestrians arrive the left boundary, they will return back from the right boundary. Thus, the total number of each type of individuals is constant and the density of pedestrian flow during each simulation experiment is fixed.

**2.2. Basic Rules.** As the walking back behavior is not considered in the model, the pedestrians can move in three different directions: those are forward, right, left (see Figure 2). The probability of moving to the neighboring grids is always not the same, depending on pedestrian walking strategy or walking habit, and so forth. For example, the probability of walking in the destination direction is the largest, which has been observed or surveyed in research work [21]. However, the probability of walking to the left or right neighboring grid is biased.

In this model, we consider that pedestrian makes decision of movement direction randomly during the walking process. Three behavior walking strategies are taken by each individual mainly according to his or her own subconscious behaviors, the distribution of other pedestrian around, and other pedestrians' walking behavior. More specifically, pedestrians are accustomed to walking along the right-hand side of the road considering the traffic rule and customs; pedestrians tend to follow other pedestrians walking behavior under the conformity psychology; pedestrians prefer to walk on the side with fewer other pedestrians there to meet his or her own personal space requirement. For convenience sake, the three walking strategies mentioned above are called *R*-strategy, *C*-strategy, and *S*-strategy, respectively, in the following.

At first, we set the transition probability of moving forward 0.70 based on the survey result in reference [21]. Then the probabilities of moving to the right and left grid are the same, while none of the walking strategies is adopted by pedestrians. When different walking strategies are taken into account, the probabilities of right and left movement are not equal and set as follows.

**2.2.1. For *R*-Strategy.** As to pedestrians with *R*-strategy, they would be used to walking on the right side. And the

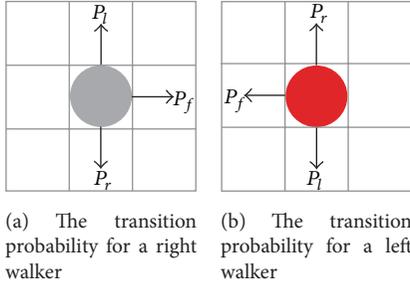


FIGURE 2: Schematic diagram of pedestrians' transition probability.

probability of moving to the right grid is higher. So the right-hand walking preference coefficient is introduced to justify the transition probability. Thus

$$p_r = \frac{K}{1+K}(1-p_f), \quad p_l = \frac{1}{1+K}(1-p_f), \quad (1)$$

where  $K$  represents the strength of right preference and  $p_f$  is the transition probability of moving forward. If one of the right and left cells is occupied by other pedestrian, the pedestrian would choose to walk to the left or right side at the probability of  $1-p_f$ . When the front cell is occupied, the pedestrian would walk from the left side or right side separately at the probability of

$$p_r = \frac{K}{1+K}, \quad p_l = \frac{1}{1+K}. \quad (2)$$

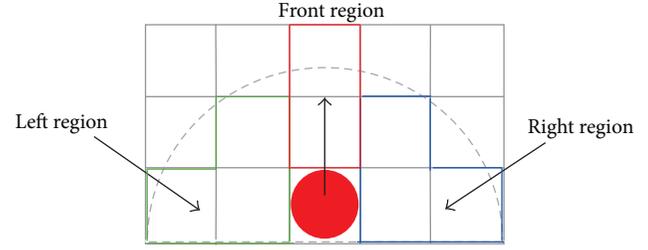
When the left (right) cell is also occupied, the pedestrian have to move to the right (left) side. And when all the three cells are occupied, the pedestrians keep still.

**2.2.2. For C-Strategy.** Those who choose C-strategy have a sense of conformity. When making the moving decision, they would take others' choice into consideration and finally follow the majority. Hence to determine the moving probability of these people, we should first collect other pedestrians' moving direction in their sight. If the majority are sure to or have the decision to walk to the left (right) side, these pedestrians will also choose to walk to the same side. Therefore, in the model, we set that pedestrians have the sight with the radius of  $r$ , and divide the sight into tripartition, the forward visual field, the left visual field and the right visual field, as shown in Figure 3. Their moving probability can be determined as follows:

$$p_r = \frac{M_r}{M_r + M_l}(1-p_f),$$

$$p_l = \frac{M_l}{M_r + M_l}(1-p_f), \quad (3)$$

where  $M_r$ ,  $M_l$  represent the number of other pedestrians in right and left visual field who chose to walk to the right or left side. And when one of the right and left cells is occupied by other pedestrians, the pedestrians would choose to walk to the left or right side at the probability of  $1-p_f$ .


 FIGURE 3: Schematic diagram of pedestrians' visual field, in the figure  $r = 2$ .

Similarly, when the cell at front is occupied, these pedestrians would walk from the left side or right side at the probability of

$$p_r = \frac{M_r}{M_r + M_l}, \quad p_l = \frac{M_l}{M_r + M_l}. \quad (4)$$

When the left or right cell is occupied, then the pedestrian would move to the opposite side. And when all the three cells are occupied, the passenger cannot move at all.

**2.2.3. For S-Strategy.** Pedestrians sometimes choose to walk on the left or right side depending on where there are less other pedestrians. Pedestrian would compare the number of other pedestrians on each visual field before they make the moving decision. Correspondingly, the probabilities of moving to right and left are

$$p_r = \frac{N_l}{N_r + N_l}(1-p_f),$$

$$p_l = \frac{N_r}{N_r + N_l}(1-p_f), \quad (5)$$

where  $N_r$ ,  $N_l$  represent the number of other pedestrians in right and left visual field. And when one of the right and left cells is occupied by other pedestrian, the pedestrian would choose to walk to the left or right side at the probability of  $1-p_f$ .

Similarly if the front cell is not empty, the pedestrian would walk to the left or right cell separately at the probability of

$$p_r = \frac{N_l}{N_r + N_l},$$

$$p_l = \frac{N_r}{N_r + N_l}. \quad (6)$$

When the left or right cell is occupied, then the pedestrian would move to the opposite side. And when all the three cells are occupied, the pedestrian cannot move at all.

According to the basic rules described above, each pedestrian can move to one of the unoccupied neighbor grids or stay at present cell at each discrete time step with certain transition probability. The update rules are applied to all pedestrians at the same time, namely, parallel update.

**2.3. Conflict Elimination.** Due to the use of parallel rule, it is possible that two or more pedestrians will choose the same target grid. Such situation is called conflicts as is shown in Figure 4. The conflicts between pedestrians are resolved by the following way [22].

- (1) Whenever two or more pedestrians try to attempt to move to the same target grid, the movement of all involved pedestrians is denied with the probability of  $u$ ; that is, all pedestrians remain at their original position.
- (2) With  $1 - u$  probability, one of the involved pedestrians is chosen randomly to move to the target grid, while others stay still.

**2.4. Repulsion Processing.** Repulsion occurs when pedestrians with different walking directions encounter. In such situation, each pedestrian will choose one side to evade. According to the experimental result, when pedestrians choose the side on which they evade, they always show a right-hand preference [23]. So the repulsion is dealt with that each pedestrian chooses to move to his or her right side, shown in Figure 5.

### 3. Simulation Result

**3.1. Parameter Definition.**  $D$  is the total density of pedestrian flow. The total number of pedestrians  $N$  is defined as the value of  $D \times W \times L$ . The average speed of pedestrians moving in one time step is defined as the value of the number of pedestrians moving forward divided by the total number of pedestrians  $N$ . The flow of pedestrians  $Q$  is defined as the value of  $D \times V$ . For each simulation, 10000 time steps were run and the first 1000 steps were discarded with the purpose of reducing error. Each simulation experiment is carried out for 10 times.

**3.2. Simulation Experiment 1: Model Validation.** Before exploring the effects of different walking strategies on bi-directional pedestrian flow, it is necessary to validate that the basic rule of model is reasonable. The simulation experiment 1 is designed for this. In this simulation experiment, walkers randomly choose one of the three strategies with the same probability during the walking process. The size of simulation system is set as follows:  $W = 30$ ;  $L = 100$ . The friction probability  $\mu$  is 0.05.

We change the density from smaller values to bigger values, and the phase transition course can be observed. As is shown in Figure 6, the red circles represent the left walker and the grey circles stand for the right walker. When the density is below critical density, the typical collective pattern of lane formation emerged for bi-directional pedestrian flow (see Figure 6(a)). While the density exceeds the critical value, the jam occurred in the system (see Figure 6(b)). These phenomena are in good agreement with empirical observations, which support the idea that the model's basic rule is reasonable.

**3.3. Simulation Experiment 2: Exploring the Effects on Critical Density.** The simulation experiment 2 is designed to study

the effect of different walking strategies on the critical density of bi-directional pedestrian flow. The width and length of system set 30 and 100, which is consistent with the simulation experiment 1. The coefficient of the right-hand walking preference  $K$  is 8 and that of sight radius  $r$  is 10. The friction probability is still 0.05.

In this simulation experiment, parts of the pedestrians in the system have certain probability to take one certain kind of walking strategies during the walking process. Then using the  $R$ -strategy,  $C$ -strategy, and  $S$ -strategy to repeat this experiment, the phenomena of critical density point drift occur. The quantitative relationship of critical density with percentage of pedestrians taking certain strategy can be occupied (see Figure 7).

It can obviously be seen that the  $C$ -curve is below the  $S$ -curve and the  $R$ -curve is at topside. What is more, the  $C$ -curve and  $S$ -curve are in declined trend with the increasing of probability, while the  $R$ -curve shows the opposite trend (see Figure 8). This means that the  $R$ -strategy has positive effect on corridor capacity, while the moment pedestrians take the  $C$ -strategy or  $S$ -strategy, the capacity reduces significantly, and jam will easily happen.

**3.4. Simulation Experiment 3: Exploring the Effects on Fundamental Diagram and Collective Phase.** Pedestrians in the corridor are set to take a same walking behavior, namely, in the three simulation, the  $R$ -strategy,  $C$ -strategy and  $S$ -strategy are made 100% in turn to study the features of different types of pedestrian flow. Although the assumed condition that all the pedestrians in the system take the same walking strategy is not existed in reality, it helps us to observe and compare the effects of different walking strategies on bi-directional pedestrian flow clearly. This is called simulation experiment 3.

The density-speed curves and density-volume curves under the condition that pedestrians fully take the  $R$ -strategy,  $S$ -strategy, or  $C$ -strategy, respectively, are drawn. It is indicated that, the pedestrian flow, no matter composed by which walking strategy, obeys the same varying tendency. There are critical density points in the curves. Below the critical density point, the speed varies gently and the volume increases with the density. When density exceeds the critical point, both the speed and volume experience a sharp decline to 0. However, it should be pointed out that there are great differences between the fundamental diagrams of the three types of pedestrian flow composed of pedestrians with different walking strategies.

By comparison, it is shown that when the pedestrian flow is not in the jamming phase, the average speed and volume of  $S$ -pedestrian flow are higher than the other two pedestrian flow under the same density. While those of  $R$ -pedestrian flow and  $C$ -pedestrian flow are basically the same when the density is no more than 0.20, yet their critical densities are different. Meanwhile,  $C$ -pedestrian flow starts to jam under the density of 0.25, while  $R$ -pedestrian and  $S$ -pedestrian flow jam under about 0.40 and 0.45, respectively.

Therefore, when pedestrians take  $S$  walking strategy, the operating efficiency of the pedestrian flow will be relatively increased, mainly indicated in the improvement of average speed, volume, and critical density. When they take  $C$  walking

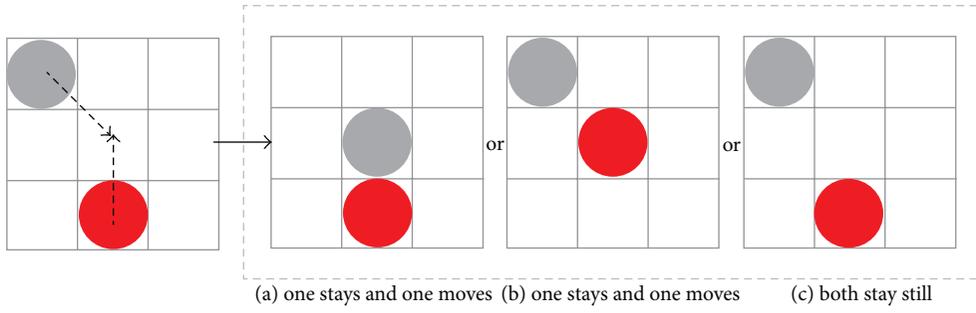


FIGURE 4: Occurrence of conflict and processing method.

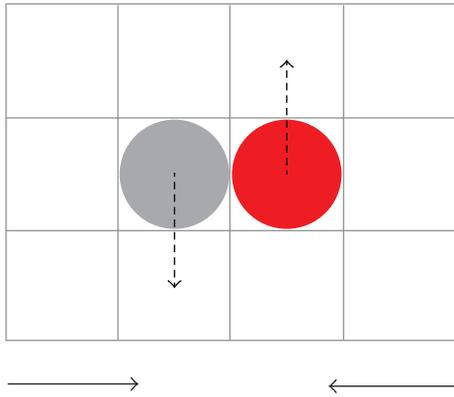
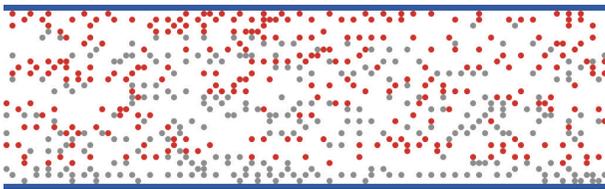
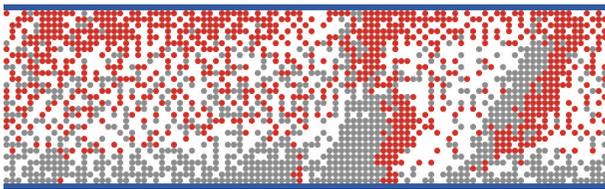


FIGURE 5: Repulsion processing.



(a) Collective phenomenon: lane formation



(b) Collective phenomenon: jam phase

FIGURE 6: Collective phenomena.

strategy, the pedestrian flow starts to jam under a smaller density. It may be explained that the herding behavior in walking leads to the gathering of pedestrians on one side or in a small area, causing congestion within one part of space. Moreover, congestion in a small or partial area is inclined to expand to the whole pedestrian flow system so as to make the system out of balance and enter the jamming phase. Therefore, it is essential to guide pedestrians with necessary information to avoid blind walking strategies taken after other pedestrians [24].

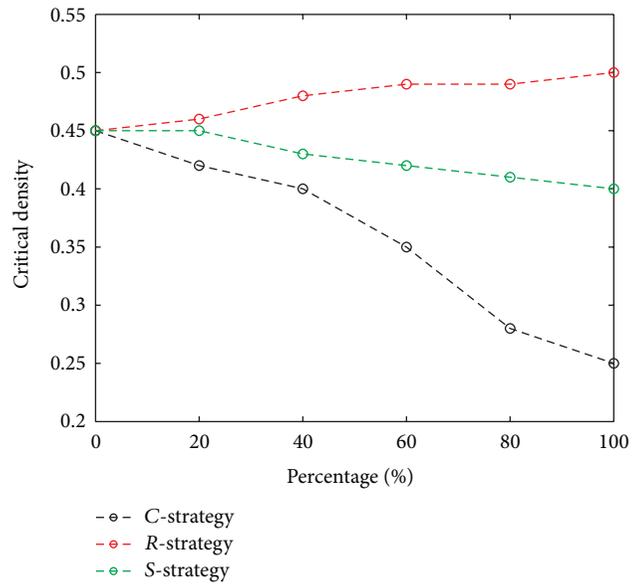


FIGURE 7: The effect of different walking strategies on critical density.

It also indicates that in the simulation, as shown in Figure 9, pedestrians with conformity psychology mainly gather in the middle of the channel while walking. While R-pedestrians are scattered to the two sides of the channel, S-pedestrians are well-distributed in the channel. When pedestrians' distribution is among a certain range, although the pedestrian density in the whole channel is low, it is very high in the pedestrian assembling area. That is why C-pedestrian flow appears to be jammed under the density of 0.25.

The definition of average speed indicates that only those who choose to move forward contribute to the average speed of the whole pedestrian flow. Though in the model, the probability of the three type pedestrians' choice of moving forward is all set to be 0.70, and the average speed differs sharply. This is mainly related to the proportion of pedestrians whose front cell is occupied by pedestrians from the opposite or same direction, and they cannot choose to go forward.

This index, therefore, is worked out, as is shown in the Figure 10. When the pedestrian density is under the critical density, the proportion of pedestrians not able to go forward increases with the density, which is also the direct cause of speed decline as pedestrian density increases. And when the density reaches the critical point, a leap occurs in

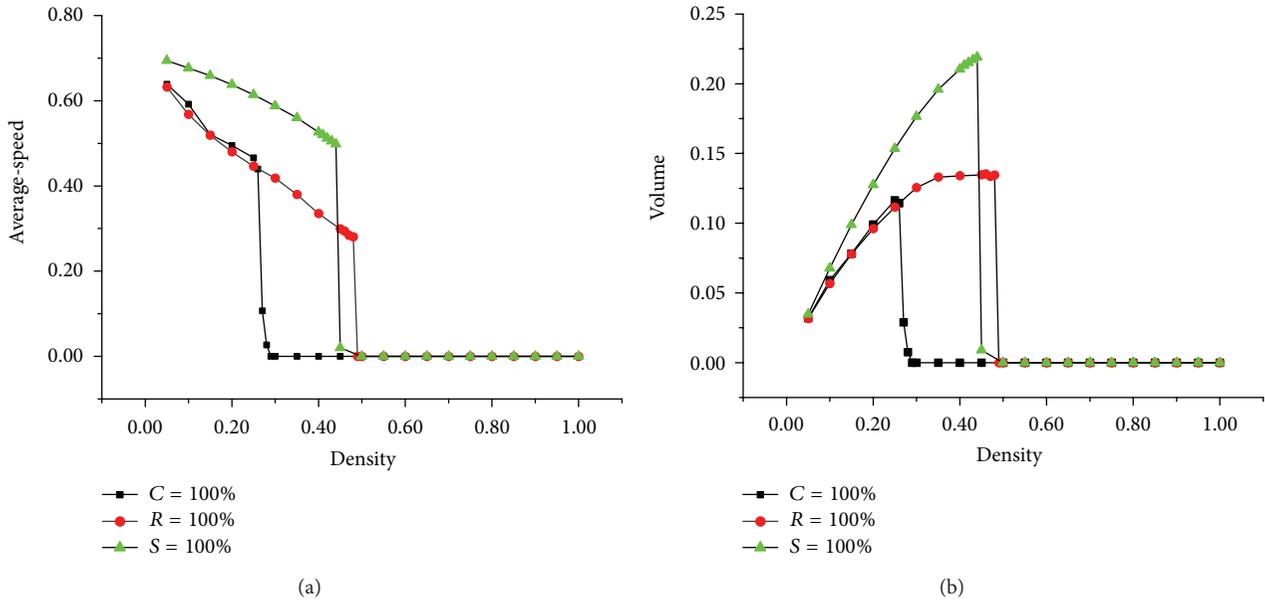


FIGURE 8: Diagram of the average speed and the volume against the density.

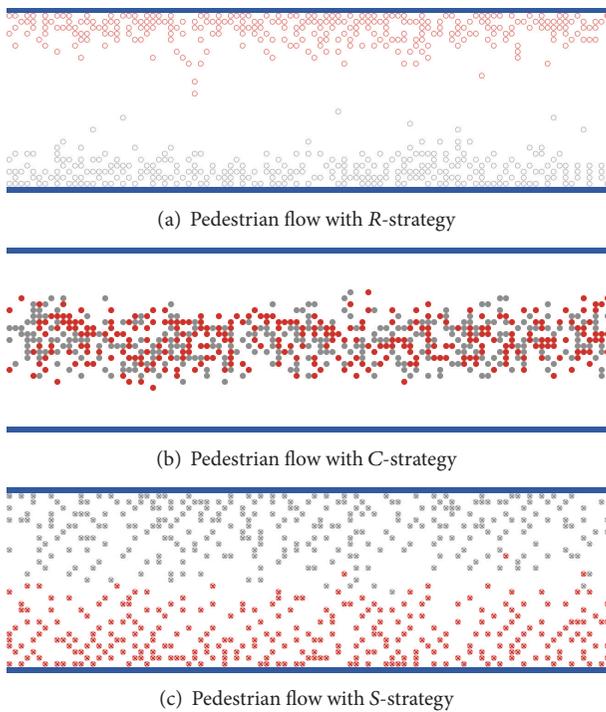


FIGURE 9: The collective phase of different pedestrian flow.

the proportion. For instance, under the density of 0.25 in C-pedestrian flow, the proportion increases suddenly from 40% to 90%. An overall comparison of the three types of pedestrian flow shows that, under the same density, the proportion not able to choose going forward is the least in S-pedestrian flow, while that in C-pedestrian flow is similar to R-pedestrian flow before it is in a jamming phase. This is because there is much space available in the channel when the density is low. Even if pedestrians begin to gather in the

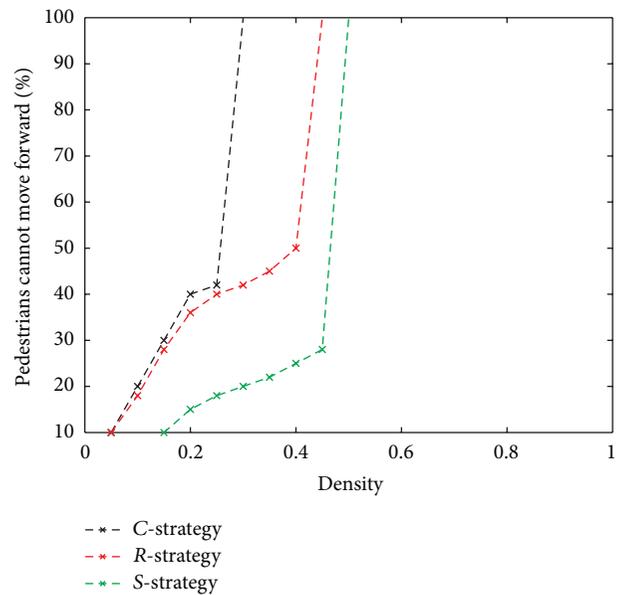


FIGURE 10: The percentage of pedestrians cannot move forward with different pedestrian flow against density.

center or on the two sides, many pedestrians' front cells in the assembling area are still empty for them to choose. However, when the density increases, instead of moving forward, most pedestrians have to stop or change steps to the left or right.

**3.5. Experiment 4: Behavior Search.** We analyze the impact of walking strategies on the pedestrian flow through research on the characters of the pedestrian flow which is purely formed pedestrians with only one walking strategy. And it is found out that the distribution of pedestrians in the channel has great impact on the movement of pedestrian flow. However, the pedestrian would not have only one type of moving

strategy because of the random combination of pedestrian crowd in reality. The real pedestrian flow is formed of pedestrians with various random moving strategies. Hence, one may expect the problem that which probabilities of  $R$ -strategy,  $C$ -strategy, and  $S$ -strategy are taken by pedestrians, respectively, when pedestrian flows have relatively high operating efficiency. If we want to explore the answer to this question, it needs to set the value of parameter  $R$ ,  $C$ , and  $S$  to simulate the pedestrian flow under every combination. Thus, according to the condition that  $C + R + S = 100\%$ ,  $C \in [0, 100]$ , and  $R \in [0, 100]$ ,  $S \in [0, 100]$ , there will be 6161 parameter combinations as the  $C$ ,  $R$ , and  $S$  are integer, which form a parameter space. Simulating every parameter combination will be a difficult and time-consuming task. To perform the search, we design the simulation experiment based on BehaviorSearch. BehaviorSearch is a software tool implemented in java and interfacing with NetLogo modeling environment which can help with automating exploration the parameter space of any models written in the NetLogo language. The desired parameters and ranges to explore, the search objective function, and the search method to be used are set as follows (see Figure 11). We run the BehaviorSearch to discover the best result when the density is 0.10, 0.20, and 0.30. The data is collected from all of the simulations running along the way, and the result is shows in Table 1.

The result of the experiment shows that the condition that makes the highest average speed of pedestrian flow is not that pedestrian flow consists of purely one type of pedestrians but mixed with  $S$ -pedestrians as majority and  $C$ -pedestrians and  $R$ -pedestrians as minority. However, when the density increases, the percentage of  $C$ -pedestrians in optimal combination will be smaller and smaller, while the percentage of  $S$ -pedestrians will be larger and larger.

Now we analyze the contour map of the average speed of pedestrian flow in the parameter space with a density of 0.20 in detail. According to Figure 12, the average speed of pedestrian flow ranges from 0.48 to 0.64, and the average speed of pedestrian flow with optimal combination is 33.3% higher than that of pedestrian flow with worst combination. This also demonstrates the fact that different walking strategy has some influence on pedestrian flow from another aspect. Meanwhile, in the contour map, speed stratification is primarily based on the percentage of taking  $C$ -strategy: when the probability of  $C$ -strategy ranges from 0% to 15%, the average speed is between 0.60 and 0.65 and when the percentage ranges from 15% to 40%, the average speed is between 0.60 and 0.56.

These results indicate that the features of pedestrian flow are related to the probability of pedestrians with different walking strategy. When the probability of taking  $C$ -strategy in the pedestrian flow is comparatively large, it will be a disadvantage to the average speed of pedestrian flow. However, if probability of taking  $C$ -strategy is controlled in a certain range (such as 15% in the simulation experiment), there will be no big influence. Another essential conclusion is that the pedestrian flow consisting of pedestrians with different walking strategies can make space resource of roads fully used to some extent, resulting in a higher average speed compared with the pedestrian flow made up of only one strategy taken by pedestrians with the same density.

TABLE 1: The results of simulation experiment based on behavior search.

Density	$R$	$C$	$S$	Max average speed
0.10	16%	11%	73%	0.695
0.20	12%	6%	82%	0.639
0.30	9%	0%	91%	0.590

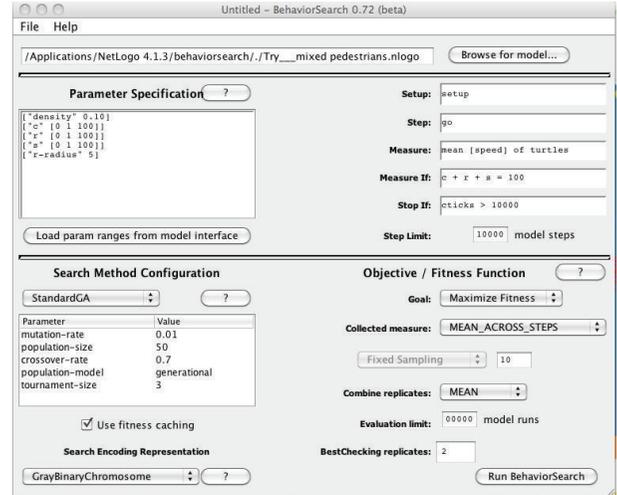


FIGURE 11: Screenshot of the BehaviorSearch GUI with the setting of parameters.

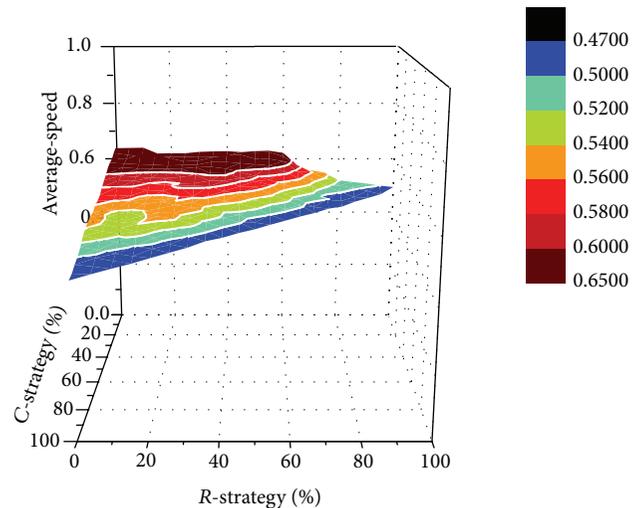


FIGURE 12: The contour map of average speed under all kinds of parameter combination.

## 4. Conclusion

Based on the research of walking habits and psychological characteristics of pedestrians, this paper summarizes three types of pedestrians' walking behavior when they are walking, namely,  $R$ -strategy,  $S$ -strategy, and  $C$ -strategy, as well as establishing the bi-directional pedestrian flow model under mixed walking strategies. In order to explore the effects on pedestrian flow of the three different walking strategies, we compared the critical density, fundamental diagrams, and collective phenomena of pedestrian flow caused by

each walking strategy. The  $R$ -strategy has positive effect on corridor capacity, while the  $C$ -strategy and  $S$ -strategy have negative one. What is more, the  $S$ -strategy contributes to improving the operation efficiency of the bi-directional pedestrian flow. Observed from the simulation, the distribution in the channel of pedestrians with different walking behavior is different. Pedestrians with  $C$ -strategy mainly gather in the middle of the channel. Pedestrians with  $S$ -strategy are uniformly distributed in the channel, while pedestrians with  $R$ -strategy gather at the two sides of the corridor. This difference of spatial distribution directly leads to the difference of the three basic diagrams of pedestrian flow. With further analysis, the reason causing speed change of pedestrian flow is the proportion change of pedestrians who cannot choose walking forward. In addition, in order to explore the problem that which is the optimal proportion of these three types of pedestrians in pedestrian flow, we designed a simulation experiment based on BehaviorSearch. According to the search of the parameter space under the condition that  $C + R + S = 100\%$ ,  $C \in [0, 100]$ ,  $R \in [0, 100]$ , and  $S \in [0, 100]$ , the optimal combination proportion when density is 0.1, 0.2, and 0.3 is obtained, respectively, and specifically analyzed the contour map of  $C$ - $R$ - $F$  against average speed when density is 0.20. For the mixed pedestrian flow with mainly  $S$ -pedestrians and a small percentage of  $C$ -pedestrians, the pedestrian flow is basically not affected. However, if the percentage of  $C$ -pedestrians in the crowd gradually increases, the average speed of pedestrian flow will decrease. These researches can help us understand the psychology and walking characteristics of pedestrians as well as the macrofeatures of pedestrian flow. They are also beneficial for traffic engineers, planners, and policy makers to develop reasonable engineering measures to guide rational distribution of pedestrians in walking space, which can be helpful to improve the operating efficiency of pedestrian flow, and avoid congestion. In our future work, the presented model will be extended to investigate pedestrian walking strategies in evacuation or crowded situation.

## Acknowledgments

This research is supported by the National High Technology Research and Development Program of China (2011AA110302) and the National Natural Science Foundation of China (no. 51078086), (51278101). The authors appreciate the support of the Jiangsu Provincial Key Laboratory of Transportation Planning and Management of Southeast University. More importantly, the authors also want to thank the interviewers and the voluntary respondents for their assistance on the data collection in the survey.

## References

- [1] D. Helbing, P. Molnár, I. J. Farkas, and K. Bolay, "Self-organizing pedestrian movement," *Environment and Planning B*, vol. 28, no. 3, pp. 361–383, 2001.
- [2] I. D. Couzin and J. Krause, "Self-organization and collective behavior in vertebrates," *Advances in the Study of Behavior*, vol. 32, pp. 1–75, 2003.
- [3] W. Daamen and S. P. Hoogendoorn, "Controlled experiments to derive walking behavior," *European Journal of Transport and Infrastructure Research*, vol. 3, no. 1, pp. 39–59, 2003.
- [4] T. Kretz, A. Grünebohm, and M. Schreckenberg, "Experimental study of pedestrian flow through a bottleneck," *Journal of Statistical Mechanics*, no. 10, Article ID P10014, 2006.
- [5] D. Helbing, L. Buzna, A. Johansson, and T. Werner, "Self-organized pedestrian crowd dynamics: experiments, simulations, and design solutions," *Transportation Science*, vol. 39, no. 1, pp. 1–24, 2005.
- [6] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995.
- [7] R. Löhner, "On the modeling of pedestrian motion," *Applied Mathematical Modelling*, vol. 34, no. 2, pp. 366–382, 2010.
- [8] Q.-M. Hu, W.-N. Fang, and Y. Deng, "Research on pedestrian movement's model based on social force," *Journal of System Simulation*, vol. 21, no. 4, article 015, 2009.
- [9] H.-H. Tian, H.-D. He, Y.-F. Wei, X. Yu, and W.-Z. Lu, "Lattice hydrodynamic model with bidirectional pedestrian flow," *Physica A*, vol. 388, no. 14, pp. 2895–2902, 2009.
- [10] V. J. Blue and J. L. Adler, "Cellular automata microsimulation for modeling bi-directional pedestrian walkways," *Transportation Research B*, vol. 35, no. 3, pp. 293–312, 2001.
- [11] F. Weifeng, Y. Lizhong, and F. Weicheng, "Simulation of bi-direction pedestrian movement using a cellular automata model," *Physica A*, vol. 321, no. 3–4, pp. 633–640, 2003.
- [12] H. Yue, H. Guan, J. Zhang, and C. Shao, "Study on bi-direction pedestrian flow using cellular automata simulation," *Physica A*, vol. 389, no. 3, pp. 527–539, 2010.
- [13] S. Sarmady, F. Haron, and A. Z. H. Talib, "Modeling groups of pedestrians in least effort crowd movements using cellular automata," in *Proceedings of the 3rd Asia International Conference on Modelling and Simulation (AMS '09)*, pp. 520–525, Bali, Indonesia, May 2009.
- [14] M. Batty, "Agent-based pedestrian modeling," *Environment and Planning B*, vol. 28, no. 3, pp. 321–326, 1993.
- [15] H. Kuang, X.-L. Li, Y.-F. Wei, T. Song, and S.-Q. Dai, "Effect of following strength on pedestrian counter flow," *Chinese Physics B*, vol. 19, no. 7, Article ID 070517, 2010.
- [16] M. Fukamachi and T. Nagatani, "Sidle effect on pedestrian counter flow," *Physica A*, vol. 377, no. 1, pp. 269–278, 2007.
- [17] S. Maniccam, "Effects of back step and update rule on congestion of mobile objects," *Physica A*, vol. 346, no. 3–4, pp. 631–650, 2005.
- [18] J. Ma, W.-G. Song, J. Zhang, S.-M. Lo, and G.-X. Liao, " $\kappa$ -Nearest-Neighbor interaction induced self-organized pedestrian counter flow," *Physica A*, vol. 389, no. 10, pp. 2101–2117, 2010.
- [19] Z. Wang, B. Song, Y. Qin, and L. Jia, "Team-moving effect in bi-direction pedestrian flow," *Physica A*, vol. 391, no. 11, pp. 3119–3128, 2012.
- [20] F. Stonedahl and U. Wilensky, "Finding forms of flocking: evolutionary search in ABM parameter-spaces," in *Multi-Agent-Based Simulation XI*, pp. 61–75. Springer, Berlin, Germany, 2011.
- [21] L. L. Lu, G. Ren, W. Wang et al., "Modeling Bi-direction pedestrian flow by cellular automata and complex networks theories," in *Proceedings of the Transportation Research Board 91st Annual Meeting*, Wash ,DC, USA, January 2012.
- [22] A. Kirchner, K. Nishinari, and A. Schadschneider, "Friction effects and clogging in a cellular automaton model for pedestrian dynamics," *Physical Review E*, vol. 67, no. 5, Article ID 056122, 10 pages, 2003.

- [23] M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G. Theraulaz, “Experimental study of the behavioural mechanisms underlying self-organization in human crowds,” *Proceedings of the Royal Society B*, vol. 276, no. 1668, pp. 2755–2762, 2009.
- [24] C. E. Dunn and D. Newton, “Optimal routes in GIS and emergency planning applications,” *Area*, vol. 24, no. 3, pp. 259–267, 1992.

## Research Article

# Optimal Skip-Stop Schedule under Mixed Traffic Conditions for Minimizing Travel Time of Passengers

Sun Feng,<sup>1,2</sup> Zhu Wen-tao,<sup>3</sup> Ye Ying,<sup>3</sup> and Wang Dian-hai<sup>3</sup>

<sup>1</sup> College of Transportation, Shandong University of Technology, Zibo 255049, China

<sup>2</sup> College of Transportation, Jilin University, Changchun 130025, China

<sup>3</sup> College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China

Correspondence should be addressed to Wang Dian-hai; wangdianhai@sohu.com

Received 24 July 2013; Revised 15 September 2013; Accepted 4 October 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Sun Feng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Given the lower efficiency resulting from the overload of bus stops, the capacity and travel time of passengers influenced by skip-stop operation are analyzed under mixed traffic conditions, and the travel time models of buses and cars are developed, respectively. This paper proposes an optimization model for designing skip-stop service that can minimize the total travel time for passengers. Genetic algorithm is adopted for finding the optimal coordination of the stopping stations of overall bus lines in an urban bus corridor. In this paper, Tian-Mu-Shan Road of Hangzhou City is taken as an example. Results show that the total travel time of all travelers becomes 7.03 percent shorter after the implementation of skip-stop operation. The optimization scheme can improve the operating efficiency of the road examined.

## 1. Introduction

In the recent years, with the rapid development of public transport, bus stops face an increasing pressure especially during peak hours, their efficiency decreases continuously, and even serious traffic congestion occurs frequently. The fundamental reason for the above problems is that bus demand exceeds the capacity of bus stops, resulting in some buses waiting in the travel lane until the buses occupying the berth entrance. This not only decreases the level of public transit service but also increases the impact between buses and cars at the location of bus stops on the road without exclusive bus lane. With skip-stop bus services that serve only a subset of stops along certain routes, this problem can be alleviated without technological improvement. For bus passengers, skip-stop services mean improved service levels in the form of lower travel time due to fewer stops and higher between-stop speed. When the skip-stop schedule is adopted by buses, it will reduce the impact of stopping buses on the cars at the location of bus stops, which will increase the car users' travel speed. In actual practice, skip-stop services in systems such as Transmilenio (Bogota, Colombia) and Metro Rapid (Los Angeles, CA, USA) have been proven to be

highly effective [1–3]. In China, Shanghai and Shenzhen have also started implementing this operating strategy. However, most cities in China use regular service which is the all-stop operational scheme considering the operational complexity and lack of methodology in modeling an optimal operational scheme.

The literature has a considerable amount of work on the transit operation optimization. Eberlein [4] formulated the stop-skipping problem as an integer nonlinear programming model with both quadratic objective functions and constraints. The real-time scheduling strategy aiming to strike an optimal balance between the benefits of operators and passengers was formulated by Fu et al. [5]. In this study, the problem is again formulated as a nonlinear 0-1 integer programming problem. Leiva et al. [3] developed an optimization method for designing skip-stop services that minimize the social costs of a segregated bus lane assuming known trip demand. An optimization method is proposed for designing skip-stop services which minimize both the travel time for passengers and the operating cost of an urban bus corridor by optimizing the transport services with different arriving frequencies of various types of buses [6]. In addition, Ceder and Wilson [7], Fan and Machemehl [8], and

Mauttone and Urquhart [9] determined a set of routes and their respective frequencies by minimizing the sum of users and transit operator costs in a given network and trip demand O-D matrix.

In those studies, the optimization methods are developed inside the transit system, while ignoring the interaction of private and transit vehicle flows. However, there are some roads without exclusive bus lanes in the cities of China. Ignoring this will lead to inaccurate estimates of travel time. Moreover, most of the previous studies were done only for a congested bus line, and they lack the analysis of the overall bus lines that passed by the stops. With these arguments as motivation, this study proposes an optimization method for designing skip-stop services that can minimize the total travel time for passengers under mixed traffic conditions by analyzing the mutual influence between buses and cars, and it uses genetic algorithm to find the optimal coordination of the stop stations of bus lines based on the objective function.

The following notations are used in describing the models in this paper:

- $C_l$  is the capacity of link;
- $C_0$  is the basic traffic capacity;
- $f_s$  is the adjustment factor for bus stops;
- $T_e$  is the impacting time by bus stops, which is determined by the number of stopping buses at this stop;
- $p$  is the probability of no bus being serviced;
- $\tau$  is the average dwell time of buses;
- $\lambda$  is the average bus arrival rate;
- $\lambda_1$  is the average arrival rate of buses that stopped at this stop.
- $S$  is the number of loading areas;
- $N$  is the number of lanes in one direction,  $N \geq 2$ ;
- $T(q)$  is the average travel time of vehicle on link  $a$ ;
- $T_0$  is the free-flow travel time on link  $a$ ;
- $q$  is the volume of traffic on link  $i$ ;
- $\alpha, \beta$  are parameters;
  - $i$  is the index of bus line,  $i = 0, 1, \dots, m$ ;
  - $I$  is the set of all bus lines in the network;
  - $A$  is the set of all links in the network;
  - $J$  is the set of all stations stopped and skipped by bus line  $i$ ;
- $q_a^{b,c}$  is the bus volume or car volume on link  $a$ ;
- $t_a^c$  is the in-vehicle travel time on link  $a$  by mode car;
- $t_{i,k}^b$  is the in-vehicle travel time on link  $k$  by mode bus;
- $n_{i,k}^b$  is the average number of passengers on bus line  $i$  on link  $k$ ;
- $n_a^c$  is the average number of passengers on car;
- $W_j$  is the total waiting time for passengers at stop  $j$ , which includes alighting and transferring at stop  $j$ ;
- $K_i$  is the set of all links where bus line  $i$  runs;

TABLE 1: Capacity of on-line linear stops (buses/h).

Number of loading areas	Dwell time (s)		
	15	30	45
1	63	43	32
2	117	80	59
3	154	105	78

Assuming 15 s clearance time, 25 percent queue probability, 60 percent coefficient of variation of dwell times, and 0.5 g/C.

$Q_{i,j}$  is the number of passengers on buses of line  $i$  at stop  $j$ ;

$\tau_{i,j}$  is the dwell time of buses of line  $i$  at stop  $j$ ;

$U_{i,j}$  is the number of passengers boarding buses of line  $i$  at stop  $j$ ;

$V_{i,j}$  is the number of passengers alighting buses of line  $i$  at stop  $j$ ;

$w_{i,j}$  is the waiting time of passengers boarding buses of line  $i$  at stop  $j$ ;

$y_{ij}$  is the decision variables to indicate stop status of bus line  $i$  at stop  $j$ , which is equal to 1 if bus line  $i$  stops at station  $j$  and 0 otherwise;

$L_j$  is the distance between station  $j$  and  $j + 1$ ;

$f_i$  is the schedule frequency of bus line  $i$ ;

$t_{oc}$  is the time of door opening and closing (s);

$a$  is passenger boarding time (s/p);

$b$  is passenger alighting time (s/p);

$\theta$  is the transfer penalty;

$x_j$  is the number of buses stopping at stop  $j$  with skip-stop operation;

$B_j$  is the capacity of stop  $j$ , from Table 1 [1].

## 2. Travel Time Model under Mixed Traffic Conditions

The travel time depends on the traffic volume and capacity of links, which is influenced by the number of stopping buses at a stop station under mixed traffic conditions. The BPR (Bureau of Public Roads) function, the most classic model, describes the link performance, which states the relationship between resistance and traffic volume [10]. In this section, the impact of bus stops on the link capacity is analyzed, and then the model of computing travel time is established based on the data of simulation runs.

**2.1. Link Capacity.** Curbside bus stops interfere with traffic flows, as the buses stop in the travel lane, resulting in a “bottleneck” (the reduction in the road width) at the location of the stops. Bus bays interfere with passing vehicles primarily, while buses maneuver to pull into and out of

the stops. In general, the capacity is calculated by adding adjustment factor for bus stops [1], as follows:

$$C_l = C_0 \cdot (1 - f_s) = C_0 \cdot \left(1 - \frac{T_e}{3600}\right). \quad (1)$$

When all passing buses stop at this stop, the impacting time is determined by the numbers of loading areas and passengers getting on and off buses. If the bus stop is considered as a queuing system, the probability of no bus being serviced can be calculated by the following equation [11]:

$$p = \left[ \sum_{k=0}^{S-1} \frac{(\lambda\tau)^k}{k!} + \frac{(\lambda\tau)^S}{S!(1 - \lambda\tau/S)} \right]^{-1}. \quad (2)$$

So the impacting time by bus stops is calculated by

$$T_e = 3600 \cdot (1 - p). \quad (3)$$

When skip-stop service is adopted, there are two possible scenarios at stops in a skip-stop operation (1) There is no bus being serviced when skip-stopped bus  $N$  arrives at this stop. (2) There is one or more buses being serviced when skip-stopped bus  $N$  arrives at this stop. In the first scenario, the bus passes directly through with no impact on the capacity of the adjacent lane. In the second scenario, the bus would pass buses being serviced on the curb lane. The lane changing behavior will affect the operation of the vehicle in the adjacent lane and result in the dropping of capacity. According to *HCM2000*, the affecting time is 4 seconds when one bus uses the adjacent lane.

The probability of the first scenario is determined by the number of buses stopped at this stop, which is computed by the following equation:

$$p_1 = \left[ \sum_{k=0}^{S-1} \frac{(\lambda_1\tau)^k}{k!} + \frac{(\lambda_1\tau)^S}{S!(1 - \lambda_1\tau/S)} \right]^{-1}. \quad (4)$$

The impacting time in a skip-stop operation is approximately calculated using

$$\begin{aligned} T_e &= 3600(1 - p_1) + 4 \times 3600(\lambda - \lambda_1)(1 - p_1) \\ &= 3600(1 + 4(\lambda - \lambda_1)) \\ &\quad \times \left(1 - \left[ \sum_{k=0}^{S-1} \frac{(\lambda_1\tau)^k}{k!} + \frac{(\lambda_1\tau)^S}{S!(1 - (\lambda_1\tau)/S)} \right]^{-1}\right). \end{aligned} \quad (5)$$

TABLE 2: Comparison of the investigated values and simulation results.

Statistical parameters	Investigated data	Simulated data
Sample size	390	683
Mean of speed	22.3	23.1
Variance of speed	21.3	34.3

Therefore, the capacity of links influenced by bus stops is calculated by

$$\begin{aligned} C_l &= C_0 \cdot \left(N - \frac{T_e}{3600}\right) \\ &= C_0 \left( N - (1 + 4(\lambda - \lambda_1)) \right. \\ &\quad \left. \times \left(1 - \left[ \sum_{k=0}^{S-1} \frac{(\lambda_1\tau)^k}{k!} + \frac{(\lambda_1\tau)^S}{S!(1 - (\lambda_1\tau)/S)} \right]^{-1}\right) \right). \end{aligned} \quad (6)$$

**2.2. Simulation Model Development and Validation.** The classic BPR function is calibrated using the traffic data of highway. Therefore, the parameters need to be recalibrated when BPR function is used to describe the performance of urban roads. It is difficult to survey traffic operating data under various conditions. So the parameters of BPR function are calibrated using the simulated data in this paper.

Based on the paper's objectives as well as the required details in the analysis, VISSIM (version 4.2) is employed in this paper. The simulation model is developed using VISSIM based on the surveyed data. For the purpose of model validation, the mean speed of values cars and buses observed on the field and simulated by VISSIM are compared. The comparison results of different types of vehicles are shown in Table 2.

A paired  $t$ -test of null hypothesis of no-mean difference is performed to check for the match between simulated and observed average speed of values vehicles. The calculated value of  $t$  is 0.6095 against the critical values (from "t" table) of 1.648. It is found that the observed and simulated average speed values agree at 5% level of significance (95% confidence limit). It can be seen that the simulated speed values significantly replicate the observed field speed for all vehicle types, and the VISSIM model can be applied to study the characteristics of traffic flow on urban road links.

**2.3. Calibration of Travel Time Model.** The BPR function is expressed as follows [10]:

$$T(q) = T_0 \left[ 1 + \alpha \left( \frac{q}{C_l} \right)^\beta \right]. \quad (7)$$

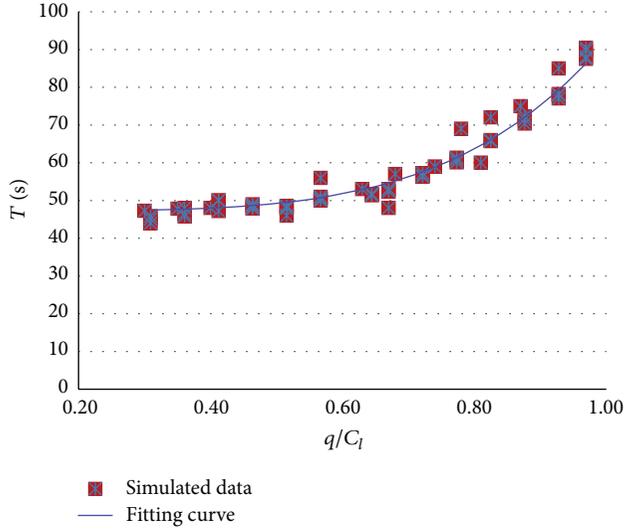


FIGURE 1: Relationship between the travel time of car and  $q/C_l$  on the segment.

In order to improve the reliability of VISSIM model, the simulation runs are made with random number seeds ranging from 41 to 45, and the average of the five values is taken as the final model output. The travel time outputted from the model is as shown in Figure 1.

In this paper, BPR model is calibrated by the method of Least-squares. First, BPR model should be deformed as a linear function by logarithmic transformation [12].

$$\ln\left(\frac{V_0}{V} - 1\right) = \ln\alpha + \beta \ln\left(\frac{q}{C_l}\right). \quad (8)$$

Denoting  $\ln(q/c)$  by  $X$ ,  $\ln(V_0/V-1)$  by  $Y$ ,  $\beta$  by  $A$ , and  $\ln\alpha$  by  $B$ , then (8) is transformed to

$$Y=AX+B. \quad (9)$$

Second, these above data are transformed to  $X$  and  $Y$ . Third, the values of  $A$  and  $B$  are estimated by the method of least-squares; then, the values  $\alpha$  and  $\beta$  are 4.2 and 0.87, respectively. Therefore, the travel time of cars is obtained when there is an on-line curbside bus stop on the road, as shown in the following:

$$T(q) = T_0 \left[ 1 + 0.87 \left( \frac{q}{C_l} \right)^{4.2} \right]. \quad (10)$$

### 3. Skip-Stop Design Model Formulation

As indicated in the previous studies, the skip-stop problem can be formulated as a nonlinear 0-1 integer programming problem, with the binary integer variables representing which stops to be skipped by the control vehicles. In this study, for the purpose of solving the exact problem, the stop-skipping problem will be formulated again as a nonlinear 0-1 integer programming problem.

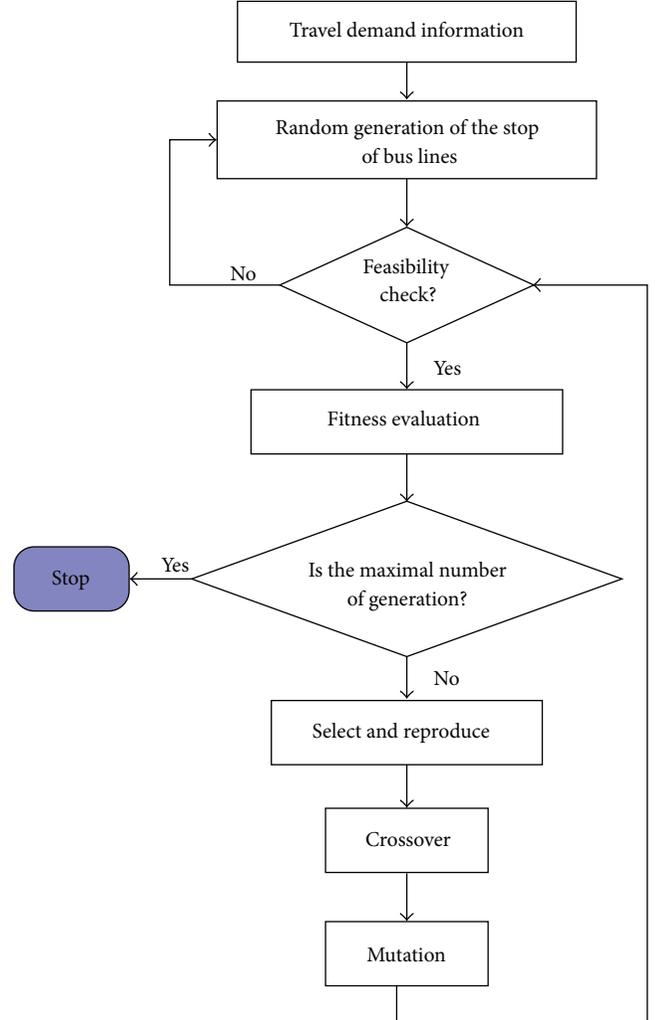


FIGURE 2: Overall procedure for finding an optimal scenario.

3.1. *Assumptions.* In order to analyze the process, the following assumptions are made in this study.

- (1) The total travel demand between any origin and destination node pair (O-D matrix) is fixed and remains the same during the analysis period. It is assumed that the travel demand is not affected by the introduction of skip-stop operation.
- (2) There are only two modes of traffic (buses and cars) utilizing the network.
- (3) Dwell time at each stop is determined by the number of passengers getting on and off the bus.
- (4) Passengers are uniformly distributed throughout the area.
- (5) A passenger would not leave the platform, and she or he would not wait more than two times.

3.2. *Skip-Stop Design Models under Mixed Traffic Conditions.* Run time models are usually used in understanding the existing service and evaluating several transit planning and

TABLE 3: Traffic data on Tian-Mu-Shan Road.

No.	Bus stop	Car volume (veh/h)	Distance between adjacent stops (m)
1	Gucui Intersection	930	—
2	Xueyuan Intersection	975	280
3	Qingfeng Village, West	893	522
4	Qingfeng Village, East	920	395
5	Xixi District	823	345
6	Bazi Bridge	760	550
7	Macheng Intersection	726	270
8	City Government	620	510
9	Hushu Intersection	723	280
10	Hangzhou Building	770	420
11	Zhongshan Intersection	780	410

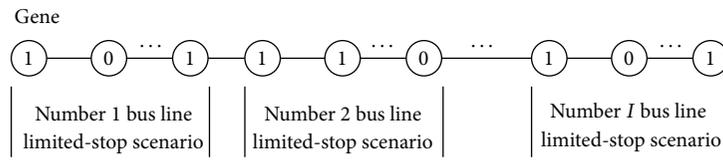


FIGURE 3: The coding illustration of a chromosome for the genetic algorithm.

operation strategies [13, 14]. So the objective function is denoted as follows:

$$\begin{aligned} \text{Min } Z = & \left( \sum_{i \in I} \sum_{k \in K_i} n_{i,k}^b t_{i,k}^b(x_j) + \sum_{j \in J} W_j \right) \\ & + \sum_{a \in A} q_a^c \times n_a^c \times t_a^c(x_j) \end{aligned} \quad (11)$$

s.t.

$$W_j = \sum_{i \in I} y_{i,j} \times (U_{i,j} \times w_{i,j} + Q_{i,j} \times \tau_{i,j}) \quad (12)$$

$$+ \theta \times (1 - y_{i,j}) \times (U_{i,j} + V_{i,j}), \quad \forall j \in J,$$

$$w_{i,j} = \theta = \frac{1}{(2 \cdot f_i)} \quad \forall i = 0, 1, \dots, n, \quad \forall j \in J, \quad (13)$$

$$\tau_{i,j} = \text{Max} \{a \cdot U_{i,j}, b \cdot V_{i,j}\} + t_{oc}, \quad \forall i = 0, 1, \dots, n, \quad \forall j \in J, \quad (14)$$

$$x_j = \sum_{i \in I} y_{i,j} \cdot f_i, \quad \forall i = 0, 1, \dots, n, \quad \forall j \in J, \quad (15)$$

$$Q_{i,j+1} = Q_{i,j} + U_{i,j} - V_{i,j+1}, \quad \forall i = 0, 1, \dots, n, \quad \forall j \in J, \quad (16)$$

$$y_{i,j} \in \{0, 1\}, \quad \forall i = 0, 1, \dots, n, \quad \forall j \in J, \quad (17)$$

$$y_{i,j} + y_{i,j+1} \geq 1, \quad \forall i = 0, 1, \dots, n, \quad \forall j \in J, \quad (18)$$

$$x_j \leq B_j, \quad \forall j \in J. \quad (19)$$

The first term in the objective function is the total travel time by buses, which includes in-vehicle travel time and total waiting time. The next term represents the total travel time

by cars. Equation (12) includes two components. The first component is the sum of waiting time of boarding and in-vehicle passengers when buses of line  $i$  make a stop  $j$ , while the second component is the transfer time of passengers with either their origin or destination stop  $j$  being skipped. Equation (13) indicates that the average waiting time and the transfer time of passengers boarding buses of line  $i$  at the stop  $j$  are equal to half the headway of bus line  $i$  assuming random arrival. Equation (14) estimates the bus dwell time at each stop based on the number of passengers who will board and alight at the stop, denoted by  $U_{ij}$  and  $V_{ij}$ , respectively. Equation (15) shows that the number of buses stopping at the stop  $j$  depends on the schedule frequency of bus line  $i$  and whether the bus line  $i$  will make stop  $j$ . Equation (16) states that the number of passengers on buses of line  $i$  at stop  $j+1$  is equal to the number of passengers on buses of line  $i$  at stop  $j$  plus the number of passengers boarding buses of line  $i$  at stop  $j$  minus the number of passengers alighting buses of line  $i$  at stop  $j+1$ . The decision variable is  $y_{ij}$  by which the system managers try to minimize their objective function ( $Z$ ). Equation (18) specifies that two adjacent stops are not both skipped by any bus line. Equation (19) ensures that the number of buses stopping at stop  $j$  cannot exceed its capacity.

#### 4. Optimization Process Using Genetic Algorithm

The proposed model is a nonlinear programming problem which associates the zero-one variables, and the parameters are tightly related to each other. It is hardly solved with conventional solution methods. The genetic algorithm (GA) is a heuristic search method that imitates the process of natural evolution [15]. It is motivated by the principles of

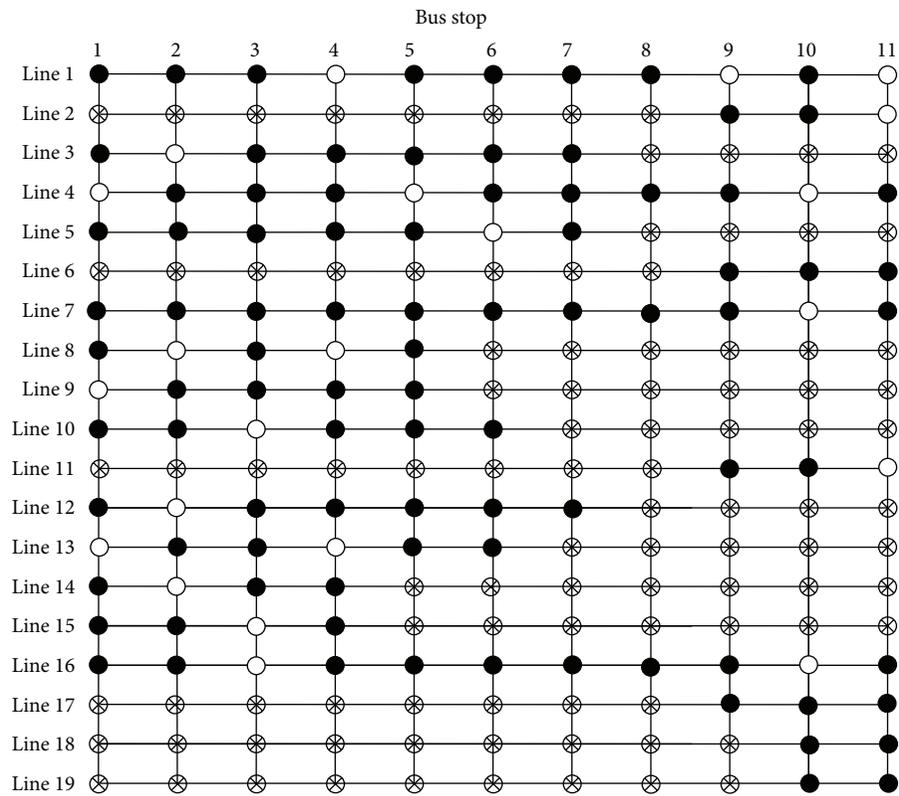


FIGURE 4: The diagram of skip-stop operation scenario. Note that “•” represents stopping, “○” represents skipping, and “⊗” represents not passing by.

TABLE 4: Numbers of passengers boarding and alighting at stops (from west to east).

Line	Stop										
	1	2	3	4	5	6	7	8	9	10	11
1	19/14	14/45	24/41	27/63	42/56	36/33	16/22	49/65	21/34	28/58	16/22
2	—	—	—	—	—	—	—	—	18/57	18/49	21/34
3	26/18	6/23	14/26	44/58	40/48	26/57	44/36	—	—	—	—
4	13/8	22/18	28/33	26/63	28/23	14/41	20/34	54/38	24/33	17/55	19/27
5	18/20	10/28	8/28	25/87	26/34	37/38	32/43	—	—	—	—
6	—	—	—	—	—	—	—	—	17/51	20/77	36/45
7	17/37	8/32	6/46	14/60	34/50	23/45	10/22	58/68	31/36	7/40	24/36
8	22/12	12/16	3/34	26/47	29/64	—	—	—	—	—	—
9	29/13	6/22	5/52	17/56	38/35	—	—	—	—	—	—
10	12/21	12/34	26/26	40/51	22/52	49/44	—	—	—	—	—
11	—	—	—	—	—	—	—	—	29/39	16/62	35/24
12	21/22	13/25	21/76	40/62	57/38	27/67	68/47	—	—	—	—
13	34/14	8/27	8/38	32/40	18/41	50/76	—	—	—	—	—
14	17/18	10/16	12/14	14/58	—	—	—	—	—	—	—
15	24/26	8/28	6/36	11/44	—	—	—	—	—	—	—
16	16/19	6/46	19/40	44/61	64/69	44/57	30/26	34/56	10/28	25/57	12/34
17	—	—	—	—	—	—	—	—	13/55	10/65	35/62
18	—	—	—	—	—	—	—	—	—	52/82	46/40
19	—	—	—	—	—	—	—	—	—	37/75	36/66

TABLE 5: The optimal scenario of skip-stop operation.

Line	Stop										
	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	0	1	1	1	1	0	1	0
2	—	—	—	—	—	—	—	—	1	1	0
3	1	0	1	1	1	1	1	—	—	—	—
4	0	1	1	1	0	1	1	1	1	0	1
5	1	1	1	1	1	0	1	—	—	—	—
6	—	—	—	—	—	—	—	—	1	1	1
7	1	1	1	1	1	1	1	1	1	0	1
8	1	0	1	0	1	—	—	—	—	—	—
9	0	1	1	1	1	—	—	—	—	—	—
10	1	1	0	1	1	1	—	—	—	—	—
11	—	—	—	—	—	—	—	—	1	1	0
12	1	0	1	1	1	1	1	—	—	—	—
13	0	1	1	0	1	1	—	—	—	—	—
14	1	0	1	1	—	—	—	—	—	—	—
15	1	1	0	1	—	—	—	—	—	—	—
16	1	1	0	1	1	1	1	1	1	0	1
17	—	—	—	—	—	—	—	—	1	1	1
18	—	—	—	—	—	—	—	—	—	1	1
19	—	—	—	—	—	—	—	—	—	1	1

Note that “1” represents stopping, “0” represents skipping, and “—” represents not passing by.

TABLE 6: The comparison between all-stop and skip-stop operations.

Scenario	Car travelers		Bus travelers		Total travel time (min)
	Travel time (min)	Out-off-vehicle time (min)	In-vehicle time (min)	Travel time (min)	
Original scenario	34,966	8,505	296,553	305,058	340,024
Optimal scenario	30,471	12,228	273,427	285,655	316,126
Change	-12.86%	43.77%	-7.80%	-6.36%	-7.03%

natural selection and survival of the fittest individuals. This method is commonly used to generate useful solutions to optimization problems [16]. Therefore, the genetic algorithm is used to solve the bilevel programming model and find the optimal coordination of the stopping stations in order to minimize inconveniences for passengers. Based on the above analysis, the procedure for finding an optimal scenario is presented, as shown in Figure 2.

For the genetic algorithm, each gene location in a chromosome represents a possible skip-stop choice, while the vehicles of bus line traverse every station shown in Figure 3, where “1” indicates stops and “0” indicates skips at the corresponding station. In Figure 3, a chromosome is divided into  $I$  independent parts. Each part indicates the corresponding operation choices for every bus line.

### 5. Case Study

5.1. *Traffic Survey.* This paper takes Tian-Mu-Shan Road (from Wan-Tang Road to Zhong-Shan Road) as an example, which is a heavy-demand corridor of Hangzhou City. The examined period is the morning peak. The traffic volume and

the number of passengers getting on and off at 11 stops are collected, as shown in Tables 3 and 4.

5.2. *Results.* Parameters values of the model and the algorithm are shown as follows: the population of chromosomes (individuals) for each generation  $N = 50$ , the crossover probability  $P_c = 0.5$ , the mutation probability  $P_m = 0.01$ , the biggest generation  $GenMax = 200$ ,  $a = b = 1 \text{ m/s}^2$ , and  $L = 500 \text{ m}$ . The optimization process is programmed using MATLAB software. Then, these data of Tables 3 and 4 are input to program, and an optimal scenario is outputted, as shown in Figure 4 and Table 5.

The travel times of all travelers before and after the optimization are compared, as shown in Table 6. It can be seen that the travel time of car travelers becomes 12.86 percent shorter than that with original all-stop operation, and the travel time of bus travelers reduces 6.36 percent due to skipping stations at the same time. The total travel time of all travelers becomes 7.03 percent shorter after the implementation of skip-stop operation. Results show that the optimization scheme can improve the overall operating efficiency of road network.

## 6. Conclusions

The flow capacity is adjusted at the section of stop zone based on the analysis of the mutual influence between buses and cars at bus stops. BPR function including the modified capacity is calibrated using traffic simulation. Then, this paper proposes an optimization model for designing skip-stop service that can minimize the total travel time of passengers under mixed traffic conditions. The genetic algorithm is used to find the optimal coordination of the stopping stations of overall bus lines in an urban bus corridor. The validation of the model and the algorithm have been proved with the help of a real-world case. Results show that the operational efficiency of buses and cars is improved with skip-stop operation under mixed traffic conditions. Although skip-stop operation is complicated for the operators and it can confuse passengers at the beginning of the service, it can certainly reduce passengers' total travel time. It should be noted that the dwell time of a bus is assumed to be based on the number of passengers boarding and alighting, as a matter of fact, which is also influenced by the number of stopping buses at the stop. Therefore, an interesting project for future research would be to propose an optimization method for designing skip-stop services with the consideration of the bus operating features at stops.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

This work is supported by the National Natural Science Foundation of China (no. 51278454).

## References

- [1] Transportation Research Board, *Highway Capacity Manual*, National Research Council, Washington, DC, USA, 2000.
- [2] Transportation Research Board, *TCRP Report 100: Transit Capacity and Quality of Service Manual*, National Research Council, Washington, DC, USA, 2nd edition, 2003.
- [3] C. Leiva, J. C. Muñoz, R. Giesen, and H. Larrain, "Design of limited-stop services for an urban bus corridor with capacity constraints," *Transportation Research B*, vol. 44, no. 10, pp. 1186–1201, 2010.
- [4] X. J. Eberlein, *Real-time control strategies in transit operations: models and analysis [Ph.D. dissertation]*, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Boston, Mass, USA, 1995.
- [5] L. Fu, Q. Liu, and P. Calamai, "Real-time optimization model for dynamic scheduling of transit operations," *Transportation Research Record*, no. 1857, pp. 48–55, 2003.
- [6] H. Niu, "Determination of the skip-stop scheduling for a congested transit line by bi-level genetic algorithm," *International Journal of Computational Intelligence Systems*, vol. 6, no. 4, pp. 1158–1167, 2011.
- [7] A. Ceder and N. H. M. Wilson, "Bus network design," *Transportation Research B*, vol. 20, no. 4, pp. 331–344, 1986.
- [8] W. Fan and R. B. Machemehl, "Using a simulated annealing algorithm to solve the transit route network design problem," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 122–132, 2006.
- [9] A. Mauttone and M. E. Urquhart, "A route set construction algorithm for the transit network design problem," *Computers and Operations Research*, vol. 36, no. 8, pp. 2440–2449, 2009.
- [10] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1985, (American).
- [11] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*, John Wiley & Sons, Hoboken, NJ, USA, 2008, (American).
- [12] M. Dongfang, W. Dianhai, B. Yiming, and S. Di, "A method of signal timing optimization for spillover dissipation in urban street networks," *Mathematical Problems in Engineering*, vol. 2013, Article ID 580546, 9 pages, 2013.
- [13] M. Mesbah, M. Sarvi, and G. Currie, "New methodology for optimizing transit priority at the network level," *Transportation Research Record*, no. 2089, pp. 93–100, 2008.
- [14] L. Y. Yao, L. S. Sun, W. H. Wang, and H. Xiong, "Adaptability analysis of service facilities in transfer subway stations," *Mathematical Problems in Engineering*, vol. 2012, Article ID 701852, 12 pages, 2012.
- [15] G. Lin, P. Liang, P. Schonfeld, and R. Larson, "Adaptive control of transit operations," Final Report for Project MD-26-7002, University of Maryland, College Park, Md, USA, 1995.
- [16] Y. Yin, "Genetic-algorithms-based approach for bilevel programming models," *Journal of Transportation Engineering*, vol. 126, no. 2, pp. 115–119, 2000.

## Research Article

# Using CVIS to Improve Bus Schedule Adherence: A Predictive Control Strategy and Its Hardware-in-the-Loop Field Tests

**Wei Yin, Jing Teng, Xiaoguang Yang, and H. Michael Zhang**

*Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, 4800 Cao'an Road, Jiading, Shanghai 201804, China*

Correspondence should be addressed to Jing Teng; [tengjing@tongji.edu.cn](mailto:tengjing@tongji.edu.cn)

Received 22 August 2013; Accepted 22 September 2013

Academic Editor: Huimin Niu

Copyright © 2013 Wei Yin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The ability of buses to adhere to their advertised schedule is vital to the bus operations. In this paper, an adaptive control strategy is proposed to dynamically adjust bus speed and traffic signal timings along the path of a running bus to improve its schedule adherence. The strategy relies on real-time location and speed information of buses provided by cooperative vehicle infrastructure system (CVIS) and uses key-time nodes calculated by back-stepping of planned arrival times to dynamically update signal timing plans to keep the bus running on time. A hardware-in-the-loop (HIL) field test was conducted to evaluate the developed strategy and the results are encouraging.

## 1. Introduction

Punctual service is vital to scheduled transit system operations but is often difficult to achieve in bus operations because its shared use of the road with other motor vehicles and the presence of traffic signals on its route. Traffic signals, bicycles and pedestrians, and uneven loading times at bus stops, to name a few, often cause buses to deviate from their advertised schedule.

Some work has been done in improving the reliability of bus arrival time by active signal control [1–3]. The primary method used in these studies is signal priority for buses [4, 5]. Bus signal priority, however, has two limitations. First, if there are many concurrent requests from different bus lines in each intersection, not all bus lines will get signal priority and their schedule cannot be guaranteed. Second, without prediction, the recognition of a bus being late may occur too late to have adequate time to restore the bus' schedule.

In order to proactively control traffic signals to support reliable bus schedules, it is desirable to predict bus arrival times in sufficient accuracy. Some complicated algorithms have been utilized to predict bus arrival times. The support vector machine (SVM) [6] and artificial neural network (ANN) [7] are two such algorithms. The data processing

mode of an ANN attempts to mimic that of a human brain, while that of the SVM is based on statistical learning theory. Both algorithms “learn” from examples provided in historical data to extract patterns for prediction. Such data were used to be obtained from inductive loops and/or other location based surveillance such as license plate readers.

With the development of vehicle-to-vehicle and vehicle-to-infrastructure technologies for improving traffic safety and traffic efficiency [8, 9], the implementation of cooperative vehicle infrastructure system (CVIS) would enable the location and velocity information of vehicles to be transmitted in real time to each other and to local traffic controllers [10, 11]. Coupled with more accurate traffic information provided by CVIS, the SVM or ANN algorithm can be used to predict bus arrival times more reliably.

In this paper, we develop a dynamic control strategy for the schedule adherence of buses under the CVIS environment and with the support of a reliable arrival time prediction algorithm such as SVC. Based on the real-time bus location and speed and the predicted arrival time under prevailing traffic conditions, it dynamically updates traffic signal timing plans at key intersections to keep the bus running on schedule. Since CVIS systems are not yet available, hardware-in-the-loop (HIL) test system is developed to evaluate the

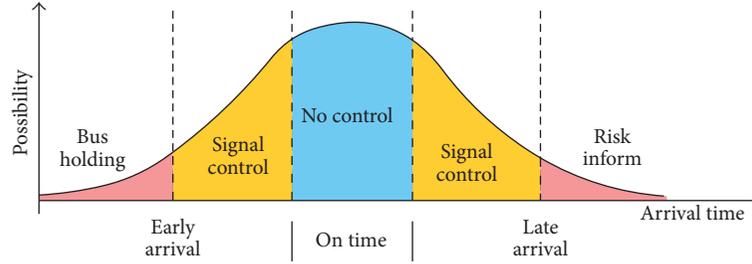


FIGURE 1: The conditions and corresponding actions.

effectiveness of the proposed control strategy. Together with field test and virtual simulator tests, HIL is often used to evaluate new ITS technologies [12–14].

The remainder of this paper is organized as follows. In Section 2, the basic concept and the logic of the control strategy are introduced. In Section 3, the HIL test platform is described and the test results are presented. Finally conclusions are given in Section 4.

## 2. The Basic Concept and Control Logic

*2.1. The Basic Concept.* Bus delays caused by the signal controls and random events can grow significantly as they travel along the path, akin to the Butterfly Effect shown in chaotic systems. For example, a small change in the system can develop into a catastrophic change over time. Similarly, a slight fluctuation in a bus' arrival time at an intersection may cause a long delay after it passes several intersections, which is one of the main causes of unpunctual bus service.

In Figure 1, a bus operating condition is divided into three categories with their possible frequencies of occurrences: early arrival, on time, and late arrival. In this figure, possible actions that help keep a bus' regular schedule are also shown. For the conditions that fall in the area marked by yellow, where bus unpunctuality is caused by random events like interference from bicycles/pedestrians and enlarged by the Butterfly Effect, using signal control alone may be sufficient to keep buses on schedule. It is in this range of conditions that our proposed bus schedule adherence control strategy applies.

The main idea of the proposed control strategy is recognizing the boundary at which the butterfly effect starts affecting the punctuality of buses and updating the signal timing plans in subsequent intersections to restore the bus' schedule. The strategy is also appropriate to the control of early bus [15, 16].

The planned arrival times are those times published in a bus schedule and the latest planned times are simply those of the planned time adding a tolerable amount of delay. Supposing that both of them have been known ahead of time, two back-stepping lines (green line and orange line) starting from the two arrival time points can then be drawn as shown in Figure 2 and the direction of the back-stepping lines is opposite to that of the bus trajectories. Those lines mark the boundaries of bus trajectories that would either adhere to schedule or be slightly late from the scheduled arrival times.

An intersection node is then defined as a key-time node if the back-stepping lines pass through its red phase. The control algorithm updates the signal timing plans of the key-time nodes to keep the bus trajectory within the bounds given by the two back-stepping lines. For example, a bus can take one of the three trajectories shown in Figure 2 (trajectories 1, 2, and 3) due to a small change of bus speed. The bus arrives at its stops either on time or within a tolerable amount of delay if it takes trajectories 1 and 2 but deviates further and further from its schedule if it takes trajectory 3. The proposed control would recognize the delay causing event at the second intersection and adjust, based on the computed back-stepping lines, the signal timing plans at subsequent intersections so as to steer the bus out of trajectory 3 and into trajectories 1 or 2.

CVIS plays an important role in the recognition of a late bus and the execution of corresponding measures because it establishes bidirectional communication between vehicles and signal controllers. Real-time bus location and speed are collected from vehicle and transmitted to control unit to predict the bus trajectory. It is then used to compare with the back-stepping lines to decide the corresponding measure. If it is speed adaption, the guide-related message is sent back to the vehicle from the control unit. And if it is signal adjustment, the message is sent to the signal controller.

The bidirectional communication based on CVIS makes the control fault-tolerant and precise. The actuated control only is unreliable because drivers' behavior is unpredictable without speed adaption. On the contrary, speed adaptation only also lacks stable effect due to the interference from the surrounding vehicles, which can be solved by actuated control. The bidirectional communication integrates two types of measures and the estimation error of travel time or the fluctuation of driver behaviors is dynamically revised in the control.

*2.2. The Operational Control Algorithm.* Before we describe the control algorithm in detail, we first list the notations used in the rest of the paper in Table 1.

The process of executing the developed control strategy is shown in Figure 3. It consists of four steps: (1) no action, (2) speed adaption, (3) signal adjustment, and (4) delay notification. It takes no action unless the current speed of the bus may fail to keep the bus running within the stable region formed by the key nodes. If it is detected that the bus arrival time at a node falls outside the region, the speed adaptation

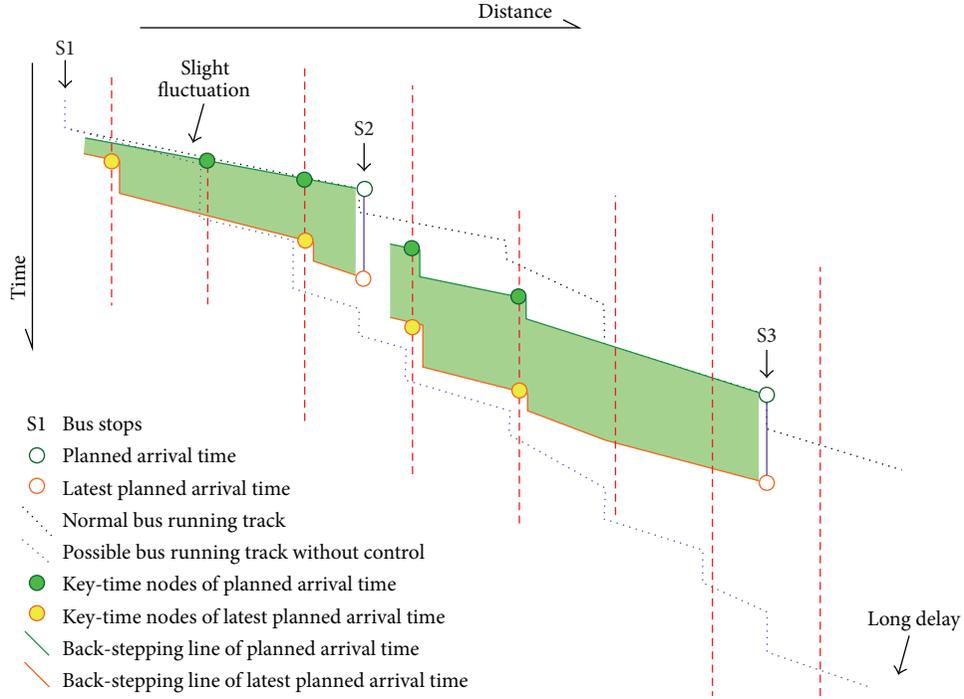


FIGURE 2: The basic concept of the proposed control.

TABLE 1: Notations.

Type	Symbol	Definition
Set	$i \in I$	The set of intersections
	$l \in L$	The set of links
	$bs \in BS$	The set of bus stops
	$tp \in TP$	The set of key-time node of planned arrival time
	$tl \in TL$	The set of key-time node of latest planned arrival time
Timing variable	$K_i$	Signal time when bus arrives at intersection $i$
	$C$	Common cycle
	$t_i^a$	Predicted arrival time point in one cycle when bus arrives at intersection $i$
	$T_i^{te}$	Red time between $t_i^a$ and the ending time of the red phase at intersection $i$
	$r_i^a$	Back-stepping arrival time point in one cycle when bus arrives at intersection $i$
	$T_i^{ts}$	Red time between $t_i^a$ and the starting time of the red phase at intersection $i$
	$t_{bs}^{TP}$	The planned arrival time point to $bs$ downstream stop
	$t_{bs}^{TL}$	The latest planned arrival time point to $bs$ downstream stop
	$t_0$	Current time
Data	$V_c$	Current speed of bus
	$V_l$	Empirical bus speed at link $l$
	$L_l$	Length of link $l$
	$L_{bs}$	Distance between bus and $bs$ th downstream stop
	$l_0$	Current link
	$l_{bs}^s$	Link in which $bs$ th downstream stop is located
	$V_l$	Empirical bus speed at link $l$

procedure is launched and the bus driver is notified. It is still possible that speed adaptation alone fails to bring the bus arrival times back within the stable region. Then the signal timing adjustment procedure kicks into the correct bus

trajectory. It checks the scope for improvement in the next several intersections downstream of the bus' current position, and if it is feasible to bring the bus trajectory inside the stable region, the control algorithm will update the signal timing

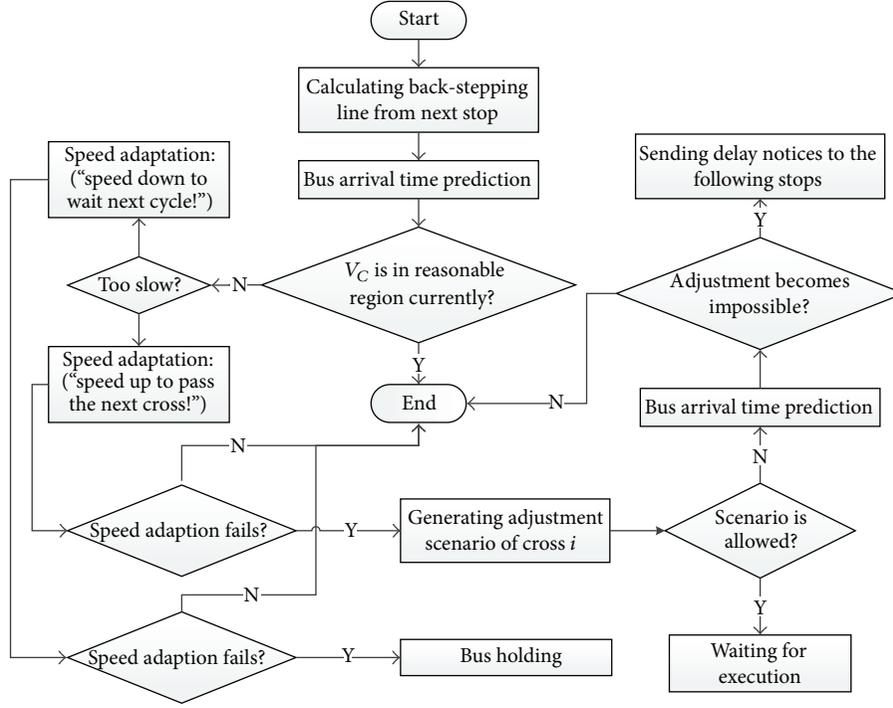


FIGURE 3: Flow chart of the control algorithm.

plans of the subsequent intersections accordingly, and if it is not possible to do so, a delay notice would be released to passengers waiting at the corresponding bus stops.

Next we describe the signal updating procedure in more detail. First, the back-stepping lines of planned arrival times are computed using the prevailing travel speeds. These are done in Steps 1 and 2. The key-time nodes are collected into the set  $\{TP\}\{TL\}$ . Next the predicted bus trajectory is obtained in Step 3 and checked against key-time nodes (Step 3-(3), (6)). If  $t_i^a > tp$  or  $t_i^a > tl$ , the control algorithm starts generating signal adjustments (Step 4), which makes either a green extension in the green phase or a red truncation in the red phase. With signal timing adjustments made, the algorithm recalculates the predicted arrival times of the bus (Step 4-(3)). If the new bus arrival times fall within the green band in Figure 2, the signal adjustments are stored and wait to be executed. Otherwise, the control algorithm moves to Step 5 and estimates the delay risk of each downstream stop (Step 5-(2), (3)).

*Step 1* (calculating the back-stepping line of planned arrival time). Consider the following:

- (1)  $t_{tmp} \leftarrow t_0^{TP}$ ;
- (2) if  $l = l_0$  then go to Step 2; else  $r_i^a \leftarrow t_{tmp} - L_l/V_l; l \leftarrow l - 1$ ;
- (3) if  $r_i^a \in \text{green phase}$  then  $t_{tmp} \leftarrow r_i^a; i \leftarrow i - 1$ ; go to (2); else go to (4);
- (4)  $r_i^a \leftarrow r_i^a - T_i^{ts}; \{TP\} \leftarrow r_i^a; t_{tmp} \leftarrow r_i^a; i \leftarrow i - 1$ ; go to (2).

*Step 2* (calculating the back-stepping line of latest planned arrival time). Consider the following:

- (1)  $t_{tmp} \leftarrow t_0^{TL}$ ;
- (2) if  $l = l_0$  then go to Step 3; else  $r_i^a \leftarrow t_{tmp} - L_l/V_l; l \leftarrow l - 1$ ;
- (3) if  $r_i^a \in \text{green phase}$  then  $t_{tmp} \leftarrow r_i^a; i \leftarrow i - 1$ ; go to (2); else go to (4);
- (4)  $r_i^a \leftarrow r_i^a - T_i^{ts}; \{TL\} \leftarrow r_i^a; t_{tmp} \leftarrow r_i^a; i \leftarrow i - 1$ ; go to (2).

*Step 3* (bus arrival time prediction). Consider the following:

- (1)  $t_{tmp} \leftarrow t_0$ ;
- (2) if  $l = l_0^s$  then go to Step 5; else go to (3) and  $t_i^a \leftarrow t_{tmp} + L_l/V_l; l \leftarrow l + 1$ ;
- (3)  $tp \leftarrow \{TP\}$  or  $tl \leftarrow \{TL\}$  //element with minimized  $i$ ;
- (4) if  $t_i^a \in \text{green phase}$  then  $t_{tmp} \leftarrow t_i^a; i \leftarrow i + 1$ ; go to (6); else go to (5);
- (5)  $t_i^a \leftarrow t_i^a + T_i^{te}; t_{tmp} \leftarrow t_i^a; i \leftarrow i + 1$ ; go to (6);
- (6) if  $t_i^a < tp$  or  $t_i^a < tl$  then go to End; else go to Step 4.

*Step 4* (signal adjustment). Consider the following:

- (1) remove  $tp$  from  $\{TP\}$  or  $TL$ ;
- (2) adjustment action at intersection  $i$ ;
- (3) recalculate  $T_i^{te}$  and  $i \leftarrow i + 1$ ; go to Step 3.

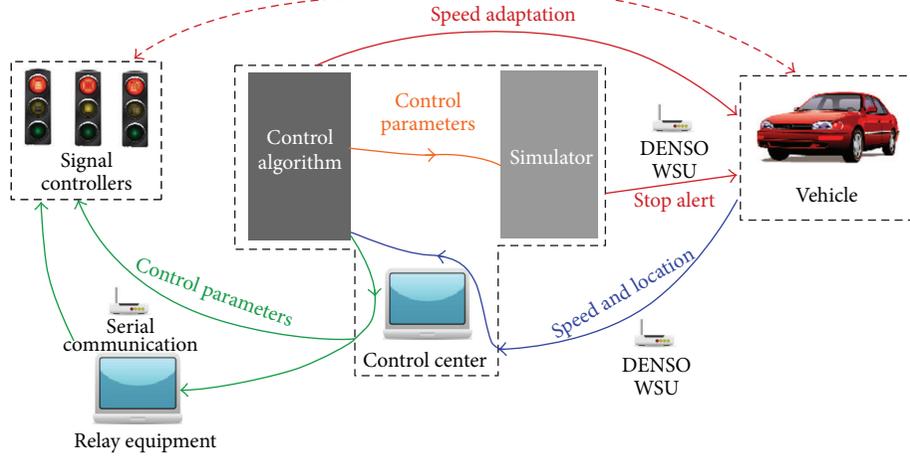


FIGURE 4: HIL simulator framework.

Step 5 (final processing). Consider the following:

- (1) if  $t_i^a < tl$  then go to End; else go to (2);
- (2) declare delay at  $bs$  stop; go to (3);
- (3)  $bs \leftarrow bs + 1$ ; if  $t_{bs}^{TL} > t_i^a + \sum_K L_i/V_i K \in \{l_0, l_{bs}^s\}$  go to End; else go to (2).

### 3. Evaluation Using HIL Tests

**3.1. The HIL Platform.** A hardware-in-the-loop simulator is used to evaluate the proposed adaptive bus schedule adherence control strategy. As shown in Figure 4, the HIL simulator consists of two field hardware components (a specific vehicle (bus) and field signal controllers) and one simulation component (a control center which consists of control algorithms to generate signal timing plan updates and a simulator to generate and move virtual vehicles). Between the three components there are four sets of communication links labeled with different colors: the control center receives the real-time vehicle status via the blue link; the center controls the traffic lights in the field through the green links and releases the speed adaptation command to the field vehicle through the red link; signal timing plan updates are transmitted to the simulator via the yellow link; an alert message from the control center is sent to the bus driver via a red link if the bus has a high risk of “crash” with virtual vehicles. Among these communication links, the simulator-to-vehicle links are realized by DENSO wireless safety unit (WSU) and the simulator-to-controller links are realized by wireless serial communication.

The HIL simulator interface is shown in Figure 5. The simulator uses a mesoscopic model in which the red dot represents the real vehicle and the black dots represent the virtual ones. Except for the virtual vehicles, all the objects in this system are real.

**3.2. Validation.** A loop in the new Tongji University campus shown in Figure 6 was selected as the test site. The loop consists of four road segments with lengths of 150 m, 250 m,

TABLE 2: Fixed signal times of the four signals in the test.

	Green phase (s)	Red phase (s)	Yellow phase (s)
Signal1	20	20	3
Signal2	25	20	3
Signal3	30	25	3
Signal4	15	15	3

145 m, and 230 m, respectively. There are four two-phase signalized intersections in the loop, whose timing plans are shown in Table 2. Two real bus stop locations were set on the longer road segments, which represent the six virtual stops. As is shown in Figure 5, the same loop is created in the simulation with virtual vehicles traveling across the intersections and on the loop. In the test, a car is used to mimic a bus and the speed characteristic scales down due to the scale differences between the loop and the real roads.

For the simulation of random interference on the road, a small application running in the test vehicle would make a sound regarding a random value (negative is allowed) at a random moment to instruct the driver to speed up or slow down. To simulate possible conflicts generated from traffic in the opposite direction and from the cross streets, the system has an option to reject signal priority requests from the test vehicle (bus).

Two sets of experiments were carried out, one without adaptive signal control and one with adaptive signal control. In each experiment, the bus vehicle runs through the loop three times and hence makes six stops. The experiment is repeated 20 times for each test set. Seven typical bus trajectories without control (purple lines) and two trajectories with control (red and pink lines) are shown in Figure 7. In the plot showing trajectories with control, the trajectories without control were also shown in the background, and the planned arrival time (marked as blue points and lines), back-stepping lines (green lines and orange lines), and corresponding actions are labeled as well.

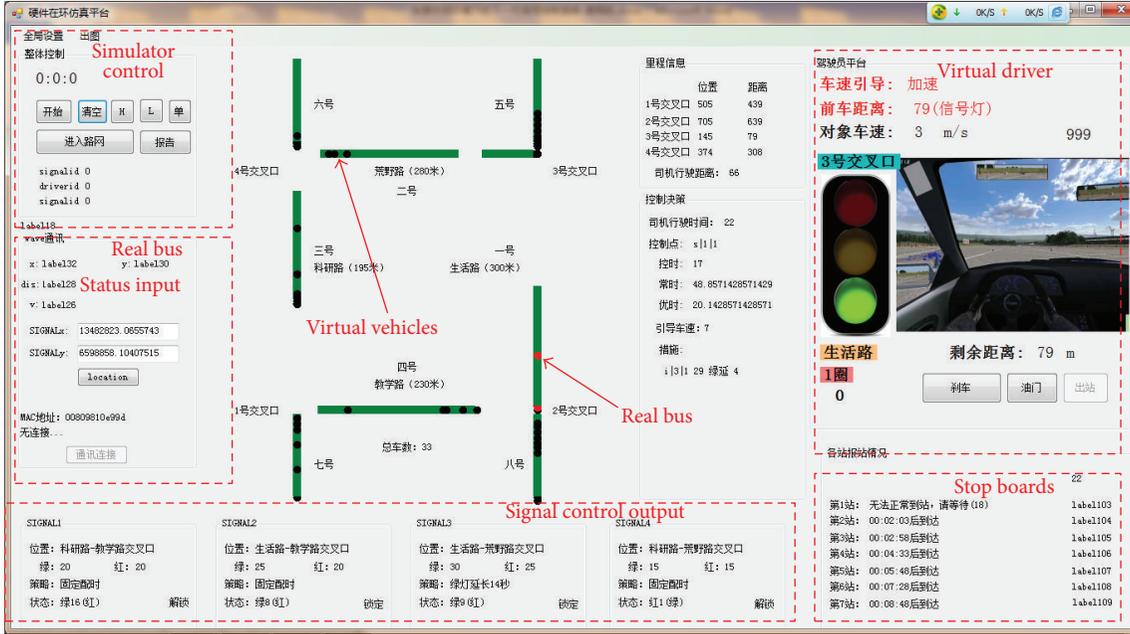


FIGURE 5: HIL simulator interface.



FIGURE 6: Field equipment and the test site.

To estimate the performance of the control strategy, the punctuality index  $PI(s)$  is created as follows:

$$PI = \frac{\sum_{i=0}^N |\text{planned arrival time}(i) - \text{actual arrival time}(i)|}{N}, \quad (1)$$

where  $N$  is the number of samples in each bus stop. The  $PI$  results of the tests are shown in Figure 8.

As shown in Figure 7, when there is no control, the bus schedule drifts away from the planned schedule further and further as the bus travels along the route, which is reflected in the  $PI$ : it rises from 5 s to 47 s in the subsequent stops (the bus made six stops in each run). The change of  $PI$  (Group 1) has close relationship with the distance between the two physical stops, the number of intersections, and the offset between signals.

The two sample trajectories presented in Group 2 in Figure 7 demonstrate the robustness of the control algorithm: the

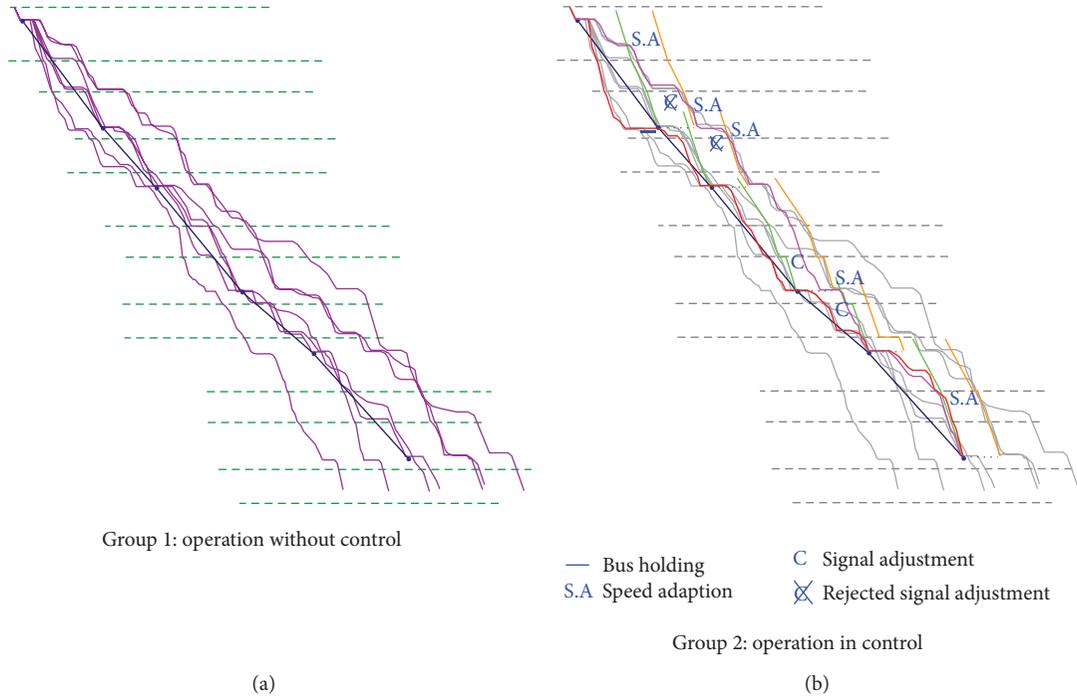


FIGURE 7: Experiment results.

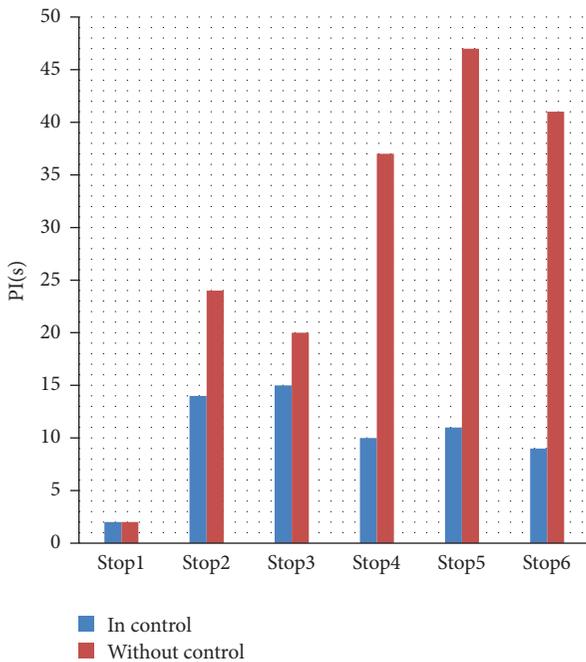


FIGURE 8: Punctuality index PI(s) in each stop.

bus received five speed adaptation messages but one of them failed to be implemented and the traffic signals were requested to change their timing plans four times but were only able to accommodate two such requests, yet the bus (pink lines) was able to run within 5 s to 15 s of the planned schedule despite these failings.

### 4. Conclusions

The proposed HIL simulation platform could reflect the performance of CVIS and the control strategy because all communication links and most objects are real. And given that there is no real crash between vehicles, the entire test is relatively more secure than the test on the real traffic system.

The HIL simulation results of the study have shown that the proposed strategy is feasible. The punctuality of the bus is effectively raised with the implement of the control strategy.

The strategy is fault-tolerant and precise due to the implement of the bidirectional communication based on CVIS. The reasonable estimation error of travel time is allowed and some failures of speed adaptation or signal adjustment have few fatal influences on the whole.

It should be noted that in the current work our primary objective is to keep buses running on schedule, and the control algorithm does not consider the potential delays to cross street traffic and traffic traveling opposite to the bus in detail. We will evaluate this potential adverse effect in our future work and refine our control strategy if such adverse effect is serious.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This research is supported by the National Natural Science Foundation of China (Project no. 61174185). The authors

wish to thank Dr. Wanjing Ma from Tongji University for his helpful suggestions and useful comments on an earlier version of the paper.

## References

- [1] P. B. Mirchandani and D. E. Lucas, "Integrated transit priority and rail/emergency preemption in real-time traffic adaptive signal control," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 8, no. 2, pp. 101–115, 2004.
- [2] N. B. Hounsell, B. P. Shrestha, F. N. McLeod, S. Palmer, T. Bowen, and J. R. Head, "Using global positioning system for bus priority in London: traffic signals close to bus stops," *IET Intelligent Transport Systems*, vol. 1, no. 2, pp. 131–137, 2007.
- [3] Y. Wadjas and P. G. Furth, "Transit signal priority along arterials using advanced detection," *Transportation Research Record*, no. 1856, pp. 220–230, 2003.
- [4] F. Dion, H. Rakha, and Y. Zhang, "Evaluation of potential transit signal priority benefits along a fixed-time signalized arterial," *Journal of Transportation Engineering*, vol. 130, no. 3, pp. 294–303, 2004.
- [5] H. R. Smith and B. Hemily, *Signal Priority (TSP): A Planning and Implementation Handbook*, 2005.
- [6] B. Yu, W. H. K. Lam, and M. L. Tam, "Bus arrival time prediction at bus stop with multiple routes," *Transportation Research Part C*, vol. 19, no. 6, pp. 1157–1170, 2011.
- [7] R. Jeong and L. R. Rilett, "Bus arrival time prediction using artificial neural network model," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (ITSC '04)*, pp. 988–993, October 2004.
- [8] H. Hartenstein and K. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164–171, 2008.
- [9] J. Anda, J. LeBrun, D. Ghosal, C. N. Chuah, and M. Zhang, "VGrid: vehicular adhoc networking and computing grid for intelligent traffic control," in *Proceedings of the 2005 IEEE 61st Vehicular Technology Conference (VTC '05)*, pp. 2905–2909, June 2005.
- [10] S. Hallé and B. Chaib-draa, "A collaborative driving system based on multiagent modelling and simulations," *Transportation Research Part C*, vol. 13, no. 4, pp. 320–345, 2005.
- [11] S. K. C. Lo, "A collaborative multi-agent message transmission mechanism in intelligent transportation system—a smart freeway example," *Information Sciences*, vol. 184, no. 1, pp. 246–265, 2012.
- [12] S. Röglinger, "A methodology for testing intersection related Vehicle-2-X applications," *Computer Networks*, vol. 55, no. 14, pp. 3154–3168, 2011.
- [13] O. Gietelink, J. Ploeg, B. de Schutter, and M. Verhaegen, "Development of advanced driver assistance systems with vehicle hardware-in-the-loop simulations," *Vehicle System Dynamics*, vol. 44, no. 7, pp. 569–590, 2006.
- [14] J. Alonso, V. Milanés, J. Pérez, E. Onieva, C. González, and T. de Pedro, "Autonomous vehicle control systems for safe crossroads," *Transportation Research Part C*, vol. 19, no. 6, pp. 1095–1110, 2011.
- [15] W. Ma, X. Yang, and Y. Liu, "Development and evaluation of a coordinated and conditional bus priority approach," *Transportation Research Record*, no. 2145, pp. 49–58, 2010.
- [16] A. Skabardonis, "Control strategies for transit priority," *Transportation Research Record*, no. 1727, pp. 20–26, 2000.

## Research Article

# Robustness Measure of China's Railway Network Topology Using Relative Entropy

Fenling Feng and Lei Wang

*School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China*

Correspondence should be addressed to Fenling Feng; [ffl0731@163.com](mailto:ffl0731@163.com)

Received 4 July 2013; Accepted 5 September 2013

Academic Editor: Wuhong Wang

Copyright © 2013 F. Feng and L. Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study focused on China's railway network topology issue from a robustness measure perspective. Relative entropy is used in the study as a measurement of robustness of railway network topology. It is found that the entropy-based measure provides more informative analyses compared to a traditional graph measure. The results indicate that the railway network in the 12th five-year plan has improved robustness when compare to 2008 with respect to deliberate and random attacks.

## 1. Introduction

Since China's first railway was built in 1865, there have been tremendous achievements in its development over 145 years. Many new stations and lines were constructed; a huge network has been formed. The planning of a railway network should not only be evaluated from external indexes, but also analyzed on its overall performance. The research on railway network topology structure takes the railway network as a complicated system and studies its topology structure using statistical methods, to provide a theoretical basis and foundation for network planning, construction, and scientific research. The robustness of network topology is one of the most important and basic features of complicated systems.

Research on robustness is based primarily on graph theory, which includes average shortest path, node degree, aggregation coefficient, and the number of interfaces. But these measures of robustness do not offer adequate amounts of information. This article uses relative entropy as the robustness measure of China's railway network topology. It also analyzes the relationship of relative entropy and ordering employing information theory methods from a systematic view. The robustness of China's railway network in 2008 and in the 12th five-year plan was analyzed.

At present, in robustness research of network topology, scholars usually take definitions from graph theory as measures of robustness. Albert et al. [1] compared

the connectivity of random networks and scale-free networks (a typical complex network) and further pointed out that the robustness of the two networks is significantly different under random attacks and malicious attacks. Watts [2] and Strogatz [3] put forward the concept of small world networks. Cohen et al. [4] found the robustness of an Internet network under random attacks and malicious attacks is based on Percolation theory. Bollobas and Riordan [5] made a mathematical analysis for robustness complex networks based on random graph theory. And Bars et al. [6] proposed a robustness measure in accordance with the Gastinel-Kahan theorem investigating stability conditions of discrete perturbed closed-loop systems. Rossi [7] presented new results for assessing the robustness of a configuration for multipurpose machines. The robustness measure of a configuration is returned by assessing the minimum magnitude of disturbances affecting the forecast demand that may lead to breaking the deadline provided by the decision maker. Border et al. [8] found that the macroscopic structure of the Web is considerably more intricate than suggested by earlier experiments on a smaller scale.

The following research used graph theory and measured the invulnerability of the network topological structure from the angle of connectivity. Holme et al. [9] used the overall efficiency  $E$  and max connected subgraph  $S$  to measure the network performance after an attack. Moreno et al. [10] found the test of strength for complex networks such as

the Internet was more stringent than others recent tests, like the random removal of nodes. Xiao and Dong [11] defined the connectivity of  $G$ , and it represents average communication paths between all nodes of the group, which is to measure and evaluate the survivability of network topology. Lu and Dong's [12] domain is the topology of the network invulnerability  $G$  as the total number of paths  $P$  which may establish within a network group divided by the pathways  $S$  between the groups. Criado et al. [13] considered it necessary for the network survivability index to have some important characteristics of the definition of network survivability. Then, they gave the invulnerability index a function definition and two kinds of survivability functions.

Railway networks have been hot topic system studies and measure methods. Banik and Dasgupta [14] deal with the use of Petri nets in modeling railway networks and designing appropriate control logic to avoid collisions. Feng et al. [15] used a spline interpolation method, a numerical differential five-point formula and the method of least squares to solve for a synergistic coefficient. Tomoeda et al. [16] presented a rescheduling method of homogenization to alleviate congestion of crowded train. Khemakhem et al. [17] provided an accurate estimate of schedule robustness, and they introduced surrogate measures. Santiago et al. [18] found through empirically analyzed that the SDH network operated by Telefónica in Spain shares remarkable topological properties with other real complex networks, such as the Internet network. Wang et al. [19, 20] analyzed the methods for the safety of traffic systems.

For years, scholars focused on using entropy to measure complex systems. Xu and Hu [21] proposed a new degree dependence entropy (DDE) descriptor to describe the degree dependence relationship and corresponding characteristics. Anand and Bianconi [22] defined the Shannon entropy of a network ensemble and proposed that it relates to the Gibbs and von Neumann entropies of network ensembles.

The paper will concentrate on the domain of proposing a new robustness approach for measuring China's railway network. A model based on a simulated China's railway network will be described. Results from an implementation with a random attack network will be reported.

## 2. Model of Railway Geography Network

From a geographical point of view, the main elements which constitute a railway network are the railway station and the railway line. Therefore, the actual railway network can be abstracted as a geography network by viewing the railway station as nodes and the railway line between each node as a side. With regard to the construction of the railway geography network, the following assumptions are needed:

- (1) the railway line is bidirectional; the railway geography network is undirected as directions of its sides are not taken into account;
- (2) the importance of the railway station and the railway line is relative; that is to say, the railway geography network is an unweighted network;

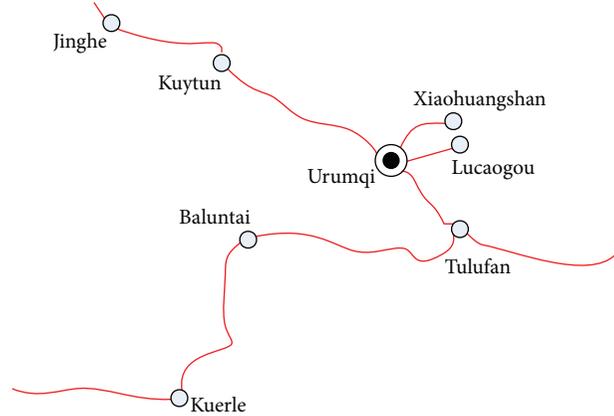


FIGURE 1: Actual Railway Network.

- (3) the carrying capacity of the railway line will not be taken into consideration;
- (4) multiple sites in the same city are regarded as a unified one (e.g., Beijing railway station, Beijing west railway station, Beijing south railway station, and Beijing north railway station in Beijing, are all seen as "Beijing").

According to this assumption, we can build a railway geography network. The network is an undirected and unweighted diagram connected by the railway station and the railway line, which can be described as  $G = (N, A)$ , of which  $N$  represents a collection of various railway stations and  $A$  represents a collection of railway lines connecting each railway station.

The following takes Urumqi station and its nearby stations as an example to illustrate this railway geographic network model. Figure 1 is the actual railway network. The railway geographic network can be worked out after it is abstracted, and a sketch can be derived. That is Figure 2, where  $G = (N, A)$ . From Figure 2, we can further obtain its adjacency matrix, shown in Figure 3.

As there are five thousand stations of different levels or so across the country, it is hard to obtain so much data. This paper researches some of the main sites according to China's railway line diagrams published by the Ministry of Railways in December 2008 and in the 12th five-year plan released in October 2012. Some main sites are to be taken to establish the 2008 railway geography network and the 12th five-year plan railway geography network.

## 3. Measure Model of Railway Geography Network

**3.1. Definition of Relative Entropy.** Entropy is a measure of the uncertainty of random variables, and it is also a measure of the amount of information required to describe random variables in an ordinary sense.

Relative entropy is a measure of the distance between two random distributions. In statistics, it corresponds to

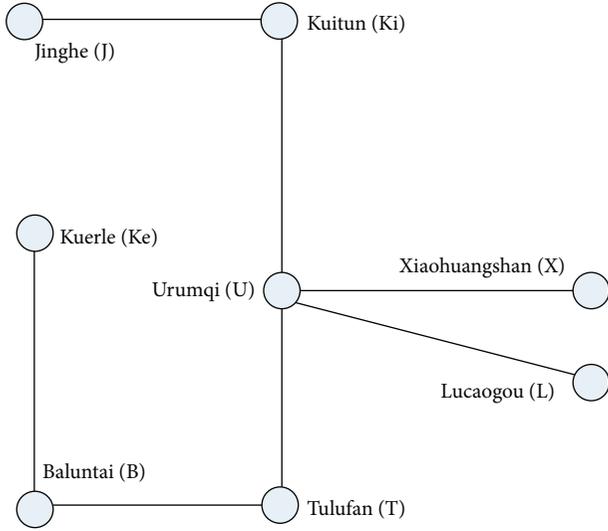


FIGURE 2: Railway Geographic Network.

	J	Ki	U	X	L	T	B	Ke
J	0	1	0	0	0	0	0	0
Ki	1	0	1	0	0	0	0	0
U	0	1	0	1	1	1	0	0
X	0	0	1	0	0	0	0	0
L	0	0	1	0	0	0	0	0
T	0	0	1	0	0	0	1	0
B	0	0	0	0	0	1	0	1
Ku	0	0	0	0	0	0	1	0

FIGURE 3: Adjacency matrix.

the logarithm expectation of likelihood ratios. The relative entropy or Kullback-Leibler distance of two probability density functions  $p(x)$  and  $q(x)$  is defined as follows:

$$D\left(\frac{p}{q}\right) = \sum_{x \in A} p(x) \log \frac{p(x)}{q(x)}. \quad (1)$$

In accordance with the intrinsic meaning of relative entropy, relative entropy can be viewed as the ordering distance between the railway network and a standard network. Therefore, as long as the completely disordered network is determined, the ordering distance between the railway network in 2008 and a completely disordered network can be obtained by relative entropy. Then we can get quantitative data on the railway network robustness. Therefore, the relative entropy can be seen as a measure of the ordering of the railway network in 2008. The robustness of the railway network can be obtained by analyzing the variation of relative entropy when the railway network is attacked.

3.2. *Properties of Relative Entropy.* In line with convention, the relative entropy possesses properties as follows:

$$0 \log \frac{0}{0} = 0, \quad 0 \log \frac{0}{q} = 0, \quad (2)$$

$$p \log \frac{p}{0} = \infty \quad (\text{based on continuity}).$$

Therefore, if there is a character  $x \in A$  to satisfy  $p(x) > 0$ ,  $q(x) = 0$ , then  $D(p/q) = \infty$ .

In addition, relative entropy has nonsymmetry; namely,  $D(p/q) \neq D(q/p)$ . The proof is as follows.

Provided that  $A = \{0, 1\}$ ,  $A$  has two distributions,  $p(0) = 1 - \xi$ ,  $p(1) = \xi$ ,  $q(0) = 1 - \delta$ ,  $q(1) = \delta$ , then the following expressions are established:

$$D\left(\frac{p}{q}\right) = (1 - \xi) \log \frac{1 - \xi}{1 - \delta} + \delta \log \frac{\xi}{\delta}, \quad (3)$$

$$D\left(\frac{q}{p}\right) = (1 - \delta) \log \frac{1 - \delta}{1 - \xi} + \delta \log \frac{\delta}{\xi}.$$

Suppose that  $\xi = 1/2$ ,  $\delta = 1/4$ , the following expressions can be obtained by calculating:

$$D\left(\frac{p}{q}\right) = \frac{1}{2} \log \frac{1/2}{3/4} + \frac{1}{2} \log \frac{1/2}{1/4} = 1 - \frac{1}{2} \log 3 \approx 0.208,$$

$$D\left(\frac{q}{p}\right) = \frac{3}{4} \log \frac{3/4}{1/2} + \frac{1}{4} \log \frac{1/4}{1/2} = \frac{3}{4} \log 3 - 1 \approx 0.189. \quad (4)$$

So it is proved.

The nonsymmetry of relative entropy refers to the distance from  $p(x)$  to  $q(x)$ , and is different from the distance from  $q(x)$  to  $p(x)$ . So, it is a must to determine what the network  $p(x)$  and  $q(x)$  represent, respectively. Typically,  $p(x)$  represents true distribution, while  $q(x)$  represents theoretical distribution. So,  $p(x)$  will be deemed the railway network distribution being studied, and  $q(x)$  will be treated as the railway network with a completely disordered distribution.

3.3. *Railway Network with a Completely Disordered Distribution—Maximum Information Entropy.* Next, the completely disordered distribution of the railway network will be determined. Information theory only gives the solve function for the maximum entropy. This paper will take China's railway network as a complex system. Topology information entropy is at its maximum when the railway network is in a uniform structure at the moment the ordering of the system is worst; when the railway network is configured with a star structure, topology information entropy is at its minimum, and here the ordering of the system is best. The construction of the railway network is a process that moves from disorder to order. Therefore, the condition when ordering is at its worst is the worst case for the railway network. This paper selects maximum information entropy as the complete railway network with a disordered distribution.

The relative entropy is the ordering distance between the railway network being studied and the railway network with

a completely disordered distribution. This means that the larger relative entropy is, the more ordered the railway network is.

When the railway network is completely disordered, all nodes have the same degree distribution  $1/n$ ; that is to say, there is  $p(x_i) = 1/n$  for any node  $x_i \in A, i = 1, 2, \dots, n$ . At the moment the maximum information entropy is the following:

$$H_{\max} = -\sum_{i=1}^n \frac{1}{n} \log_2 \frac{1}{n} = -n \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n. \quad (5)$$

**3.4. Measure of Railway Geography Network Robustness.** Accepting the definition of relative entropy as a measure of robustness requires a distribution as a basis. This paper is intended to employ node degree distribution.

Degree is defined as the number of adjacent sides of the nodes; namely  $k_i = \sum_j a_{ij} = \sum_j a_{ji}$ , which represents the total number of the sides connected to  $i$  points. Degree distribution function  $P(x_i)$  represents the probability that any node's degree is  $k$ . In another words, degree distribution means the proportion of nodes with degree of  $k$  among the total number of nodes in the network. As an important geometrical property of the network, degree distribution could describe some characteristics of the network.

If the degree distribution density function of the railway network is expressed as  $p(x)$ , then the robustness measure  $C$  can be obtained from the expression (1) as follows:

$$C = D\left(\frac{p}{q}\right) = \sum_{x \in A} p(x) \log \frac{p(x)}{q(x)} = \sum_{x \in A} p(x) \log(np(x)). \quad (6)$$

## 4. Proof of Measure Testability of China Railway Geography Network

To define a new measure, it is necessary to analyze whether or not it is a measurable function. Next, analysis is to be conducted on whether the relative entropy function is a measurable function or not by using the Lebesgue measure theory in the function of read variable.

**4.1. Definition and Theorem.** First, a few basic definitions and theorems.

**Definition 1** (measurable set). Set  $E \subset R$ ; if any collection  $A \subseteq R$ , there is

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^C). \quad (7)$$

Then  $E$  is Lebesgue measurable or measurable in short,  $m^*$  is the outer measure,  $E^C$  is supplementary set of  $E$ , and  $\mu$  is the whole measurable set.

**Definition 2** (measurable function). Set  $E$  is a measurable set, and set  $f$  is a function from  $E$  to  $R$ . If any section  $I \subset R$ , there is

$$f^{-1}(I) = \{x : x \in E, f(x) \in I\} \in \mu. \quad (8)$$

Then  $f$  is Lebesgue measurable, or measurable in short.

**Definition 3** (measure). Measure is a set of functions  $m(\cdot)$ , defined in the collection of sets constituted by the subsets of given set  $\Omega$ , valued in the expansion of real functions  $R \cup \{\pm\infty\}$  and meeting certain conditions. These conditions should be consistent with the following properties.

- (1) Normative:  $m(\Phi) = 0$ .
- (2) Nonnegativity: for all  $E \in \mu, m(E) \geq 0$ .
- (3) Countable additivity: for any non overlapping row  $\{E_k\}_{k=1}^{\infty}$  in  $\mu$ , there is  $m(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} m(E_k)$ .

**Theorem 4.** Set  $(X, F)$  is a measurable space, and set  $f : X \rightarrow R^*$  is a function defined in the measurable space. Then the fact that  $f$  is a measurable function and the fact that for any  $B \in \bar{B}(R^1), f^{-1}(B) \in F$  is equal.

### 4.2. Construction of Measurable Sets

**Step 1.** Since the relative entropy is actually a probability distribution function of a set of random variables, a probability space  $(\Omega, F, P)$  is constructed first. When it comes to the Chinese railway network topology, the probability space can be built based on the node degree distribution. Assuming the selection of degree is the event  $(\omega_i, i = 1, 2, \dots, n)$ ,  $n$  is maximum degree of nodes, sample space is as follows:

$$\Omega = (\omega_1, \omega_2, \dots, \omega_n). \quad (9)$$

The value of degree is derived from a function  $X(\omega)$  on the sample space. From this, we can use the subsets of the sample space to make a  $\sigma$ -domain  $F$ . Proportion of degree  $\omega_i$  is the probability that a node's degree is  $X(\omega_i)$ . For this reason, set  $P(\{\omega_i\}) = P_i, i = 1, 2, \dots, n$ ; if  $A \in F(A \subset \Omega)$ , then  $P(A) = \sum_{\omega_i \in A} P(\{\omega_i\})$ . So a probability space  $(\Omega, F, P)$  is set for random phenomena as the degree distribution of Chinese railway network topology.

**Step 2.** According to the probability distribution and the definition of probability measure, a new probability space  $(R, \bar{B}, P_X)$  is induced in the above random variables, of which  $B \in \bar{B}$  ( $\bar{B}$  refers to Borel collection); there is  $P_X(B) = P(X^{-1}(B))$ . Function  $P_X(\cdot)$  is a probability distribution of random variable  $X$  and distribution in short. According to the real variable theorem, this set function  $P_X(\cdot)$  is a probability measure on the set functions  $\bar{B}$ .

Here, it is required to know that  $A$  and  $B$  are both the set of events, and the difference between them is that  $A$  refers to setting a few values for node degrees, while  $B$  refers to when the value of several node degrees are real numbers. In other words,  $A$  is a simple event and  $B$  belongs to the set of real numbers. However, all real numbers are basic events, and all Borel sets are events; hence,  $B$  is also an event.

Thus, a new probability space  $(R, \bar{B}, P_X)$  is induced from the probability space  $(\Omega, F, P)$ .  $P_X(\cdot)$  represents a probability measure of event  $B$ .

**Step 3.** As the relative entropy is the probability distribution function of the random variable, the relative entropy function  $C = \sum_{x \in A} p(x) \log(np(x))$  can be written as  $C = C(P_X)$ .

According to  $P_X(B) = P(X^{-1}(B))$ , we can figure out that  $D = C(P_X(P))$  is a composite function.

**4.3. Proof of Testability.** Next, we will prove that the relative entropy is a measurable function.

Because the expression  $C = \sum_{x \in A} p(x) \log(np(x))$  is composed of simple functions,  $C(P_X)$  is a continuous function in  $R^1$ . Because  $B, \{P_X : C(P_X) < B\}$  is an open set in  $R^n$  for any real number  $B$  and an open set is a Borel set,  $C(P_X)$  is Borel measurable. Since  $P_X$  is measurable, according to Theorem 4, there is  $P_X^{-1}(C^{-1}(B)) \in F$ . So for any  $B \in \bar{B}(R^1)$ ,  $C^{-1}(B) \in \bar{B}(R^1)$ , composite function  $D^{-1}(B) = P_X^{-1}(C^{-1}(B)) \in F$ ,  $D(P)$  is a measurable function in  $(\Omega, F)$ , according to Theorem 4.

At this point, it is proven that the relative entropy is a measurable function.

## 5. Analysis of Railway Network Robustness in “The 12th Five-Year Plan”

To test and verify the utility of the new robustness measure, two networks were selected. The first one is a circuit diagram of China’s railway network in 2008, in which 313 nodes and 472 sides were selected; the other is a circuit diagram of China’s railway network in “the 12th five-year plan”, in which there are 420 nodes and 753 sides selected (Figure 4). By comparing the relative entropy changes of railway network in “the 12th five-year Plan” and the railway network in 2008 under two different types of attacks, we can get to know each configuration’s robustness and their pros and cons.

The superiority of the relative entropy measure can be shown by comparing relative entropy with the clustering coefficient of graph theory. Clustering coefficient is defined as follows: if the node  $i$  is collected to other nodes by  $k_i$  sides, then there are  $k_i(k_i - 1)/2$  sides at most between the  $k_i$  sides. So, the clustering coefficient  $C_i$  is a ratio of the actual number of sides between the  $k_i$  sides and the maximum number of possible sides  $S_i$ .  $C_i$  can be described as  $C_i = 2S_i/k_i(k_i - 1)$ .

**5.1. Initial State.** We can determine that the relative entropy of China’s railway network in 2008 is 0.6501 and in “the 12th five-year plan” is 0.6931 by inputting the initial data of the railway geography network in 2008 and the “12th five-year plan” to the expression (6). The calculated average clustering coefficient in 2008 is 0.0699 and in “the 12th five-year plan” is 0.0918. From this, ordering and aggregation degrees of node in “the 12th five-year plan” are higher than in 2008. It is observed that the conclusion derived from relative entropy and average clustering coefficient is consistent.

**5.2. Deliberate Attack.** In this part, we’ll analyze the robustness of the railway geography network in 2008 and “the 12th five-year plan” under deliberate attack. Deliberate attack refers to those attacks on important sites on purpose. It shows up in the deletion of node degrees from largest to smallest on the geography network model. The result is shown in Figure 5.

From the point of view of relative entropy, similarities of the two railway networks under deliberate attack are as

follows. (1) Along with increasing attack strength, the relative entropy of railway network declines, and the ordering gets lower and lower. (2) The fact that relative entropy fluctuates is because that deliberate attack is from larger node degree to smaller node degree. When the nodes of large value are deleted, the railway network will be divided into several smaller networks. The relative entropy is a superposition of ordering of these small networks, so there will be a rally. (3) When deliberate attacks increase to a certain extent, the railway will be completely disordered. And when the node degree is 1, each node will be of equal status, and then the network will remain in a completely disordered state until the attack is over.

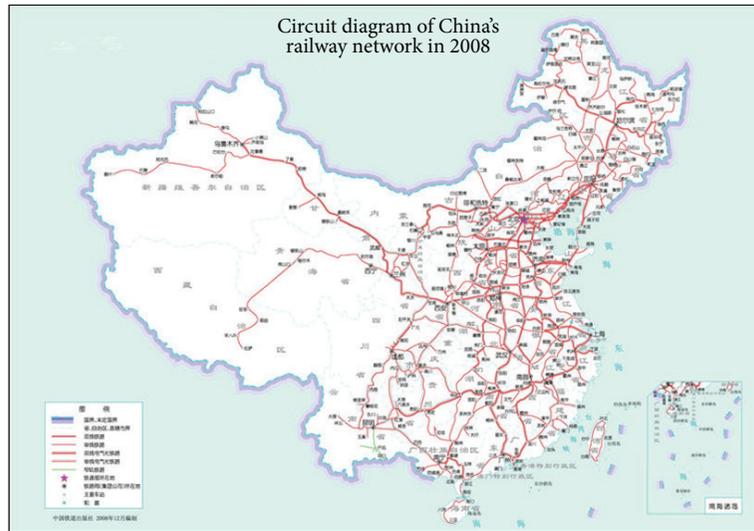
Compared to the railway network in 2008, the railway network in “the 12th five-year plan” has the following advantages. (1) During the early stage when a deliberate attack is about 10 percent, the relative entropy is decreasing slower. But the railway network in “the 12th five-year plan” still has higher ordering and stronger robustness when the deliberate attack is relatively small. (2) The change in relative entropy of the railway network in “the 12th five-year plan” is more stable than that in 2008. (3) The railway network in “the 12th five-year plan” becomes quickly disordered when the deliberate attack is at 70 percent. But the railway network in 2008 becomes quickly disordered when the deliberate attack is at 50 percent. (4) The railway network in “the 12th five-year plan” becomes completely disordered when the deliberate attack is at 90 percent. But the railway network in 2008 becomes completely disordered when the deliberate attack is 80 percent.

From the point of view of the average degree of aggregation, the average degree of aggregation in 2008 and “the 12th five-year plan” is declining under deliberate attack. When attacks reach a certain extent, the clustering coefficient is 0. The decreasing trend of average degree of aggregation in “the 12th five-year plan” is more stable than that in 2008. Clustering coefficient is more durable than that in 2008.

**5.3. Random Attack.** After a description of a deliberate attack, analysis will be given on the robustness of railway geography network in 2008 and “the 12th five-year plan” under random attack. Random attack refers to aimless attacks on sites. In the geography network model, a random attack deletes node degrees randomly. The result is shown in Figure 6.

From the point of view of relative entropy, both the railway networks in 2008 and “the 12th five-year plan” have common downward trends and volatility changes and end in complete disorder, just the same as with those under deliberate attack.

In the meantime, the railway network in “the 12th five-year plan” under random attacks still has the following advantages. (1) When a random attack is less than 10 percent, the ordering of the railway network in “the 12th five-year plan” is higher than that in 2008. (2) The railway network in “the 12th five-year plan” is more stable than that in 2008 with stronger robustness. (3) The railway network in “the 12th five-year plan” becomes quickly disordered when random attacks reach 90 percent. But the railway network in 2008 becomes quickly disordered when random attacks are at 80 percent or

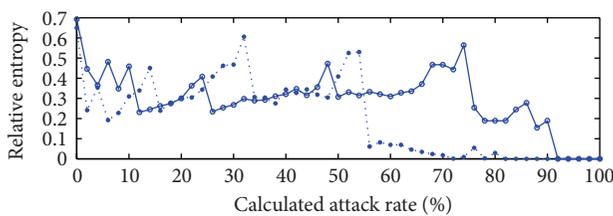


(a)

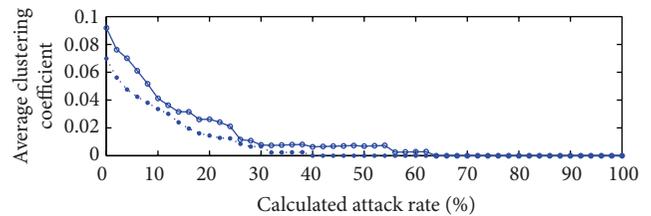


(b)

FIGURE 4: Circuit diagram of China's railway network in 2008 and in the "12th five-year Plan."



(a)



(b)

—○— Railway network in "the 12th five-year plan"  
 ···· Railway network in 2008

—○— Railway network in "the 12th five-year plan"  
 ···· Railway network in 2008

FIGURE 5: Relative entropy and average clustering coefficient change of railway network under deliberate attack.

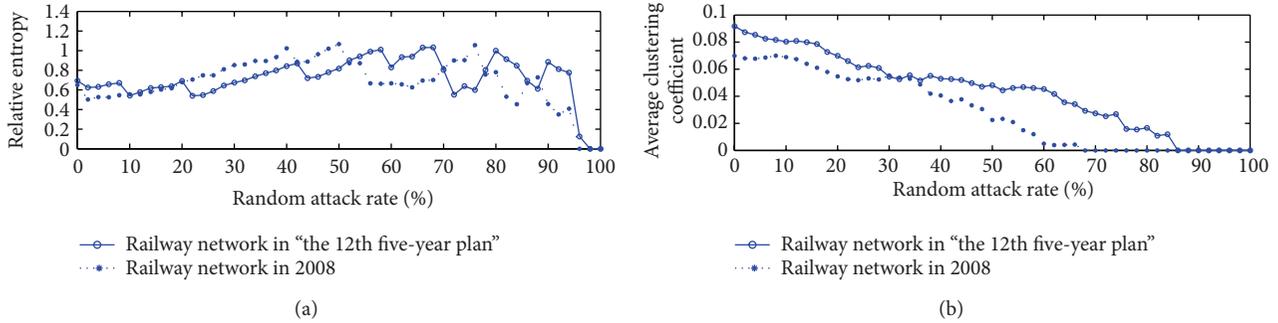


FIGURE 6: Relative entropy and average clustering coefficient change of railway network under random attack.

so. (4) The railway network in “the 12th five-year plan” arrives at complete disorder when the random attack is 98 percent. But the railway network in 2008 becomes completely disordered when the random attack is at 95 percent.

From the point of view of the average degree of aggregation, the average degree of aggregation in 2008 and “the 12th five-year plan” declines under random attack. The clustering coefficient will be 0 in the end, which is the same as that under deliberate attack.

The decreasing trend of average degree of aggregation in “the 12th five-year plan” is more stable than that in 2008. The clustering coefficient becomes 0 later than that in 2008.

By comparing the conclusions above, we can see that on the one hand, as a robustness measure, relative entropy could be used to get similar conclusions. On the other hand, the relative entropy measure can provide this information as follows. (1) The overall trend of networks under attack. (2) The rate of measure change at the initial attack stage. (3) When networks will become in rapidly disorder at the initial attack stage. (4) When networks will be in a state of almost complete disorder at the last stage. The common graph theory measure can only show the first and the fourth information. So, taking relative entropy into account can provide more accurate figures and greater amount of information.

Furthermore, representing the order of networks, the relative entropy can be used to analyze the robustness for a view of systems, which means that relative entropy has a measure function beyond graph theory.

## 6. Conclusions

This paper deals with the robustness measure approach to railway network. First, the entropy-based robustness measure approach is established. It was found that the entropy-based method provides greater amounts of information than the graph theory measure. Based on the method proposed in this paper, stronger robustness was found for the 12th five-year plan railway network as compared to that of 2008 network, with respect to deliberate attack. Furthermore, the 12th five-year plan railway network has greater stability and ordering than the 2008 network.

Further study can be conducted to compare the relative entropy measure to other graph theory and existing research on the robustness of railway traffic networks and transport networks in China.

Network robustness is not unique to railway networks, the relative entropy measure can also be used with other types of complex networks, such as biological networks, internet networks, and social networks. This method to assess the relative entropy is also universal.

## Acknowledgments

The authors would like to thank Matt Rolph and center for communication practices at RPI for the discussions and comments on the paper. Project supported by the Fundamental Research Funds for the Central Universities (Grant no. 2010QZZD021) and the Ministry of Railway Science and Technology Research Development Program (Grant no. 2012X012-A), China.

## References

- [1] R. Albert, H. Jeong, and A. L. Barabasi, “Error and attack tolerance of complex networks,” *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.
- [2] D. J. Watts, “The “new” science of networks,” *Annual Review of Sociology*, vol. 30, pp. 243–270, 2004.
- [3] S. H. Strogatz, “Exploring complex networks,” *Nature*, vol. 410, no. 6825, pp. 268–276, 2001.
- [4] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, “Resilience of the Internet to random breakdowns,” *Physical Review Letters*, vol. 85, no. 21, pp. 4626–4628, 2000.
- [5] B. Bollobas and O. Riordan, “Robustness and vulnerability of scale-free random graphs,” *Internet Mathematics*, vol. 1, no. 1, pp. 1–35, 2003.
- [6] R. Bars, I. Jaworska, and S. Tzafestas, “A mimo discrete-system robustness measure for internal model control structure,” *Systems Science*, vol. 23, no. 1, pp. 25–35, 1997.
- [7] A. Rossi, “A robustness measure of the configuration of multi-purpose machines,” *International Journal of Production Research*, vol. 48, no. 4, pp. 1013–1033, 2010.
- [8] A. Broder, R. Kumar, F. Maghoul et al., “Graph structure in the Web,” *Computer Networks*, vol. 33, no. 1–6, pp. 309–320, 2000.
- [9] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, “Attack vulnerability of complex networks,” *Physical Review E*, vol. 65, no. 5, Article ID 056109, 14 pages, 2002.
- [10] Y. Moreno, J. B. Gómez, and A. F. Pacheco, “Instability of scale-free networks under node-breaking avalanches,” *Europhysics Letters*, vol. 58, no. 4, pp. 630–636, 2002.

- [11] Z. Xiao and Z. Dong, "Improved GIB synchronization method for OFDM system," *IEEE Telecommunications*, vol. 2, no. 8, pp. 1417–1421, 2003.
- [12] H. Lu and Z. Dong, "Carrier frequency offset estimation of DAB receiver based on phase reference symbol," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 1, pp. 127–130, 2000.
- [13] R. Criado, J. Flores, B. Hernández-Bermejo, J. Pello, and M. Romance, "Effective measurement of network vulnerability under random and intentional attacks," *Journal of Mathematical Modelling and Algorithms*, vol. 4, no. 3, pp. 307–316, 2005.
- [14] M. Banik and R. Dasgupta, "Modeling of railway network using petri nets," in *Proceedings of the International Conference on Modeling, Simulation & Visualization Methods*, pp. 175–179, 2011.
- [15] F. Feng, D. Lan, and L. Yang, "Analysis on the synergy evolutionary development of the collecting, distributing, and transporting system of railway heavy haul transportation," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 329120, 22 pages, 2012.
- [16] A. Tomoeda, M. Komatsu, I. Y. Yoo et al., "Real-time railway network simulation and alleviating congestion of crowded trains," in *Proceedings of the ICROS-SICE International Joint Conference (ICCAS-SICE '09)*, pp. 4932–4936, August 2009.
- [17] M. A. Khemakhem and H. Chtourou, "Efficient robustness measures for the resource-constrained project scheduling problem," *International Journal of Industrial and Systems Engineering*, vol. 14, no. 2, pp. 245–267, 2013.
- [18] A. Santiago, J. P. Cardenas, M. L. Mouronte, V. Feliu, and R. M. Benito, "Modeling the topology of SDH Networks," *International Journal of Modern Physics C*, vol. 19, no. 12, pp. 1809–1820, 2008.
- [19] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [20] W. Wang, X. Jiang, S. Xia, and Q. Cao, "Incident tree model and incident tree analysis method for quantified risk assessment: an in-depth accident study in traffic operation," *Safety Science*, vol. 48, no. 10, pp. 1248–1262, 2010.
- [21] X. Xu and X. Hu, "Degree dependence entropy: a new descriptor for complex networks," in *Communications in Computer and Information Science*, vol. 3 of *AsiaSim*, pp. 19–26, 2012.
- [22] K. Anand and G. Bianconi, "Entropy measures for networks: toward an information theory of complex topologies," *Physical Review E*, vol. 80, no. 4, Article ID 045102, 2009.

## Review Article

# A Review Study on Traction Energy Saving of Rail Transport

**Xuesong Feng, Hanxiao Zhang, Yong Ding, Zhili Liu, Hongqin Peng, and Bin Xu**

*MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing 100044, China*

Correspondence should be addressed to Xuesong Feng; [charles.x.feng@gmail.com](mailto:charles.x.feng@gmail.com)

Received 4 July 2013; Revised 3 August 2013; Accepted 21 August 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Xuesong Feng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The energy cost of a rail transport system is very big especially for the tractions of trains in its daily operation. Continual effort on decreasing the traction energy cost intensities of various rail transport modes has been made by many researchers and practitioners on different aspects for a long time. From the rail transport operation perspective, this study reviews such energy-saving research mainly focusing on the optimizations of train coasting schemes, the rational designs and utilizations of track alignments, the ameliorations of train attributes, and the improvements of system operations. According to the review work, it is confirmed that in sound responses to distinct track alignments, dynamically optimizing control programs of trains with their reasonably improved attributes ought to be further studied in view of the systematic transport operation of a rail line or network in an integrated manner as much as possible in the future research on traction energy saving of rail transport.

## 1. Introduction

Although rail is commonly regarded as an energy-saving travel pattern, the operation of a rail transport system consumes in fact a huge amount of energy every day. For example, the weekly electric cost of merely one railway station of Hong Kong Mass Transit Railway Corporation reaches 230 megawatt hours [1]. Such huge energy consumption sometimes may have a serious energy waste. For instance, it is astonishing that the energy cost (EC) per passenger trip of a transport completed by the metro system in New York is even much higher than the EC of the same trip by car on average, mainly due to the low utilizations of the passenger capacities of the metro trains [2].

Therefore, only rationally utilizing a rail transport mode is able to avoid unnecessary EC to achieve its sustainable operation and development. For example, optimizing the streamline design of a train [3], increasing its passenger capacity utilization rate to some reasonable levels [4, 5], coasting the train, that is, in other words, taking advantage of its inertia motion, as much as possible in its transport process [5, 6], and so on all in practice take much positive effect on decreasing its EC per unit transport. In addition, traction power innovations are also able to not only effectually

improve the EC efficiency of a train [7] but also substantially reduce its emissions of  $\text{NO}_x$ , CO, PM, and so forth [8].

## 2. Energy Cost Factors

The EC of a rail transport system for its daily operation consists of two parts, that is, the traction EC (TEC) for the tractions of its trains and the additional EC for, for example, lightening in stations, repairing cars in depot(s), and so forth. The share of the TEC is ordinarily much more than half of the total EC of a rail transport system. As shown in Figure 1, the TEC intensity (i.e., the TEC per unit transport) of a train for its completion of a transport task is dynamically determined by many factors on the aspects of driving tactics (i.e., mainly the coasting program) of the train, alignments (e.g., the slope gradients) of the track, attributes (e.g., the start-stop frequency) of the train, operation management (e.g., the train scheduling) of the rail transport system, natural environment (e.g., the wind [9]) and others (e.g., the aerodynamic forces especially in a tunnel [10]) in an integrated manner.

Focusing on the impacts of various factors, many researchers and practitioners have been continually for a long time making effort for the decrease of the TEC intensity of a train or, afterwards in a further, multitrains operating on

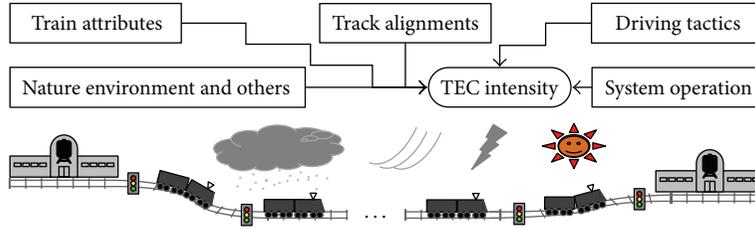


FIGURE 1: Influence on TEC intensity.

a rail transport line or network and have proposed many valuable suggestions. This study reviews these works from the transport operation perspective mainly with respect to train coasting control optimization, reasonable track alignment design and utilization, train attribute amelioration, and rail transport system operation management improvement. Accordingly, the direction of our effort in future research is indicated for radically decreasing the TEC intensity of the rail transport system operation in an effective way.

### 3. Coasting Control

A train consumes energy in a relatively very slight intensity when it coasts. Rational site choices of a train for the starts of its coasting along a rail line are able to effectively decrease the TEC of the train for its whole trip by avoiding its brakes for changing its speed with much loss of its kinetic energy. As illustrated in Figure 2, much importance in research has been attached to the optimization of the coasting scheme of a train for its energy saving in regard to transport time expenditure (TTE), train traction performance, passenger ride comfort, and so forth.

Based on train traction calculations [11, 12], a transport simulation program following the principle of coasting vector (i.e., coasting speed and coasting start point) to achieve the optimum automatic train operation with the least TEC for a trip in a certain travel time is developed by Chui et al. [13] for Kowloon-Canton Railway (KCR). A train is given a coasting vector data from the trackside so that whenever the train has actually reached the coasting speeds or coasting start points determined with a view to passenger travel demands in different daily operation time of a rail line, its motoring power will be shut off. In this way, about 3% of the TEC could be saved for KCR.

For the sake of improving the application result veracity of a train coasting program, Chang and Sim [14] newly propose a genetic algorithm (GA) [15] applied in their rail transport simulation study which takes moving block signals [16] to insure the safe distance between two neighboring trains running on the same rail line. Different sites of a transport section between neighboring stops and various control actions of a train are respectively regarded as genes and chromosomes in the GA to search for the optimal coasting start points of the train. Decreasing TEC, ensuring on-time running, and avoiding jerky actions (i.e., in other words, providing comfortable transport service to the passengers) of a train are

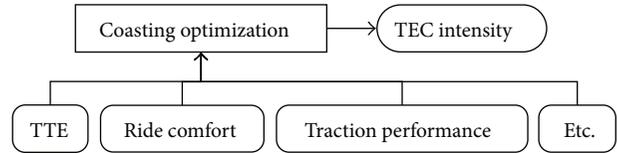


FIGURE 2: Coasting optimization for decreasing TEC intensity.

considered in a comprehensive way to optimize its driving tactics for a transport mission.

Because of the computational complexities of GAs for the exploration of the optimal coasting program in comparison with, for example, gradient algorithms [17] especially when the stop spacing of a train is comparatively very short [18], Hwang [19] tries to simplify the process of optimizing coasting controls of a high-speed railway (HSR) train for its energy saving. The relationships between the economic speed of a HSR train and its TEC and TTE for a passenger transport work are built based on fuzzy clustering analyses [20, 21]. As a result, it is possible to get the range of the economic speed of a train according to expected TEC and TTE. Thereafter, the coasting program of a train is resolved by taking advantage of a hybrid scheme of GAs to optimize the speeds of the train at different rail sites within its economic speed range.

Due to low efficiencies of classic numerical optimization methods [22] and burdensome train traction calculations in detailed simulations of the transport controls of a train [5], it is difficult to truly realize the optimum on-board controls for the minimal TEC. On the basis of existing studies, Bai et al. [23] develop a set of train operation guidance equipments to successively give suggestions to the driver of a train in its transport process for optimizing the driving actions for traction energy saving to some dynamically real-time extent. Liu and Golovitcher [22] provide an analytical approach solved by maximum principles [24, 25] for relatively efficient decisions of optimal control change points of a train adopting full or partial braking, coasting, or motoring with partial or full power. Moreover, according to the statistical distribution of the operating conditions described by the speeds and accelerations of a train on the macrolevel, Lindgreen and Sorenson [5] simplify the transport simulation work for the versatile and flexible control strategy (including coasting start points) optimization with an acceptable TEC computation accuracy.

Based on previous research on the effect of optimizing the coasting plan of a train on its TEC intensity, analyses of

the additional impacts of traction acceleration and braking capacity of a train are made by Bocharnikov et al. [26] in simulation with utilizing dynamics theories [27], GAs, and fuzzy mathematics (FM) [28]. Different coasting programs of a train for the same transport work are comparatively studied with focusing on the changes of its TTE as well as TEC. The coasting scheme is finally determined to prevent the unilateral optimization of the TEC at overmuch expense of the TTE. The utilized GA is able to effectively avoid improperly adopting local optimal solutions in searching for the most energy-saving coasting start points for the whole trip. Moreover, variables applied in the dynamic simulations are well defined by FM methods.

With regard to transport time schedule adherence, Kim and Chien [29] newly develop a train performance simulation approach to optimize the controls of a train for distinct types of track alignments to decrease its TEC for a certain trip. Under some transport time constraints for a transport distance between neighboring stops, the most energy-saving control scheme of a train for the transport section is determined according to the optimal selection of its coasting start points. In this respect, simulated annealing approach [30] is utilized to analyze the TECs of different operations of the train for various track alignments in the transport section which is divided into many small subsections. Further, Kim and Chien [31] improve their work with the additional consideration of the formation of a train. The sensitivity analyses of different factors such as track alignment, speed limit and train formation of a train are also made for its energy saving.

#### 4. Track Alignments

As previously-clarified from the viewpoint of optimizing the coasting scheme of a train, various types of track alignments have different influence upon the TEC intensity of the same train control action. Therefore, by effectively transforming potential energy of ramps along a rail line [32–34] and increasing running straightness of a train in plane [35], reasonable designs and/or utilizations of the track alignments are much beneficial to reducing the TEC intensity of a train. Valuable studies have been made on this aspect in view of, for example, driving strategy of a train, headway of trains, and ride comfort of passengers, as interpreted in Figure 3.

Focusing on updating the ramps in stations along a metro line for different transport distances between neighboring stops, Hoang et al. [36] first attempt to simultaneously optimize vertical track alignments and train operation control for traction energy saving. Different patterns of the ramps are defined according to realities. Heuristic algorithms [37] and dynamic programming (DP) [38] are used to detect the slope change points of each type of the ramps. Thereafter, a simulation approach is applied in accordance with the principles of train traction calculations to compute the TEC of a train for its trip from one terminal station to another of a rail line. The designs of the energy-saving slopes of the ramps in each of the stations and the corresponding actions of a train on different rail sites are able to be optimized rationally.

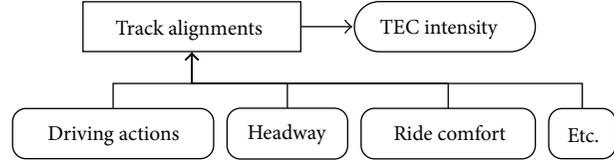


FIGURE 3: Track alignment design and utilization for reducing TEC waste.

From a unique viewpoint, Firpo and Savio [39] do research on how to reduce the electric energy provided by a transformer substation to a transport section of an HSR line for the transports of a certain number of trains within some time. The impacts of the track alignments together with the site of the transformer substation along the rail line and the headway of the trains in the transport section on the electricity supplied by the transformer substation are studied comprehensively. With the help of electromechanical system modeling approach [40], the established nonlinear optimization model is solved by both Simplex method [41] and stochastic hill-climbing algorithm [42, 43] to explore the optimal combination of target speed, acceleration, and time in stop of a train for the transport section.

On the assumption that the climbing ability of a metro train is capable of coping with all kinds of ramps, Kim and Schonfeld [44] study simulation of the influence of various types of vertical track alignments upon the intensities of both TEC and TTE of the train in consideration of the ride comfort requirements of the passengers. It is concluded that if a stop spacing is shorter than about 2,400 meters, a rational slope design of a ramp in station is able to take obviously positive effect on saving the TEC of a train for this stop spacing mainly because of its coasting after the ramp. Moreover, reducing the acceleration of a train is of benefit to decrease its TEC for the same transport work, but meanwhile, its TTE increases in correspondence.

Regarding impacts of different radians of various types of track curves in plane on the TEC intensity of a train, Liu et al. [35] make simulation analyses with employing train traction calculations. It is demonstrated that the increase of the TEC intensity of a metro train with the decrease of the track curve radius is obviously accelerated when the curve radius becomes smaller than about 300 meters. In contrast, if radiuses of track curves are over approximately 500 meters, their radians have almost no influence upon the TEC intensity of a metro train.

It is true that the speed adjustment action of the driver is lagged behind the change of the speed (control reminding) code of a train when the track gradient, speed restriction, and so forth are various between successive stops. As a result, Ke et al. [45] propose a combinatorial optimization model to minimize the TECs of a train for each of its transport sections. A MAX-MIN ant system [46] is used to optimize the train-speed trajectory for reducing the computational burden on the block-layout design. Further, the optimal speed codes of a train are also determinable for its transport in different sections to provide an energy-saving, efficient, and

comfortable service [47]. In order to insure the practicability of corresponding driving strategies, the relationships between the acceleration and speed of the train and the track gradient are regulated with a fuzzy process [48, 49] to reflect the correlations among track alignments, train actions, and corresponding TECs.

## 5. Train Attributes

Besides the optimum driving controls of a train and the rational designs and utilizations of track alignments, ameliorated attributes of a train also play important roles in its traction energy saving. Different from the viewpoints of aforementioned studies, many works pay much attention to for example, motor flux, weight, mass distribution, electricity supply effectiveness, traction power output, startup-stop frequency, formation, carrying capacity utilization rate, and so forth of a train for the improvement of its TEC efficiency. As shown in Figure 4, such research involves different fields including electrical control, rail car manufacturing, rail transport operation management, mechanics, and electric power supply.

From the perspective of regenerative braking of a train, Kokotovic and Singh [50] as early as in the 1970s have proposed an optimization approach to provide the basis for the improvement of train control circuit through optimizing the electric motor flux control to reduce the energy cost intensity of the traction motor of a train. With referring to the correlations between the speeds and traction forces of a train for different track alignments [11], the energy-saving transport objective of the train is achieved by utilizing a Hamiltonian system [51] with adjoint analyses [52] to rationally adjust the speed of the train through traction motor flux control. In addition, the relationship between the TEC and TTE of the train is also studied with the application of Green's theorem [53] for further optimization of the motor flux control.

According to train traction calculations applied in simulations of the transport procedures of different types of Chinese metro trains, it is empirically confirmed that the TEC intensity of a metro train in China is increased with raising its weight in an approximately linear way [35]. Reducing the weight of a train is a direct means to decrease its TEC intensity through reducing the resistances to the train in its trip. On the basis of the detailed survey on various technical specifications of different types of trains in Germany, IFEU [54] analyzes the general energy-saving effect of reducing 10% of the weight of each type of the trains. It is revealed that the vast majority of the TEC of an urban or regional rail transit train for its whole trip is used to accelerate the train in its startups because of its comparatively very short stop spacing on average. As a result, if the weight of such a train decreases by 10%, it is able to save about 7% or 8% of its TEC for a trip in Germany. In comparison, an intercity train with a relatively long stop spacing needs to spend roughly half of its TEC in overcoming the air resistance to its moving after a startup, and therefore, the TEC saving by lighting the weight of the train is from 4% to 5%.

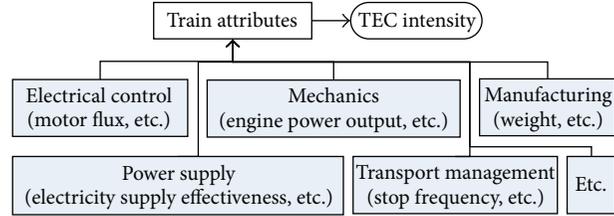


FIGURE 4: Attribute improvement for TEC intensity reduction.

As for simulation research on rail transports, initially a mass point or afterward a line with its mass evenly distributed is commonly used to roughly represent a train. In contrast, Chou and Xia [55] analyze the impact of the mass difference of various parts of a freight train on the TEC and TTE of its whole trip with regard to transport security. In their simulation work, a heavy haul train is divided into a certain number of units according to the traction power distribution of the train. Every two neighboring units is separated with protective equipments controlled by electric signals in an associated way to dynamically adjust actions of different units confronted with different track alignments. Optimal control strategies to reduce the TEC and TTE of a transport and improve its security are suggested on the basis of the simulation analyses.

For the purpose of decreasing the TEC intensity of a diesel multiple-unit train, Lu et al. [56] evaluate different combinations of the traction power outputs from the multi-engines of the train for a certain total traction power output. DP has been applied to analyze the TECs produced by a single-train motion simulator to identify the global optimal program of the traction power distribution between different engines of a train for a certain transport whose profile must be obtained in advance. To realize the online power management [57], an adaptive online strategy for the real-time optimization based on the results from the DP study is put forward. It is indicated that the optimized solution has a TEC reduction of around 7% in comparison with even distribution of the total traction power output of a train to each of its engines for the same transport mission.

A high startup stop or accelerating-braking frequency of a train substantially increases its TEC for the same transport task [5, 6]. For example, an express train running without any stops from Copenhagen to Elsinore is able to save about 50% of its TEC in comparison with its transport with every stop for this trip [5]. Feng [58] and Feng et al. [59], respectively, prove that the increase of the TEC intensity of an HSR or metro train with shortening the stop spacing is apparently accelerated when the stop spacing becomes shorter than correspondingly around 100 kilometers or 1,800 meters, and such a trend becomes more obvious if the target speed of the train is improved. In consideration of short distances between neighboring stations of most metro lines in reality, a relatively low target speed of a metro train [35, 59] and its frequent coasting operations are able to effectively decrease its TEC intensity [6, 60] especially for relatively low boarding rates in off-peak hours when the TEC intensity of the metro system

is much further decreased if the trains extend their headways [13].

As widely recognized, it is comparatively easier to make the utilization rates of the passenger capacities of shortly formed trains relatively high according to dynamically changed travel demands of the passengers, which will obviously decrease the total TEC intensity of the trains running on a rail line or network [31]. Moreover, the decrease of traction power of a train is able to reduce its TEC intensity with more TTE [26, 44]. In addition, Feng et al. [61] also reveal that in comparison with the TEC of a short HSR train with the same equipments, an HSR train with a long formation may have more TEC for the same trip due to the speed restrictions of the exit switches of the tracks in stations. All these facts require systematically reasonable and flexible applications of trains with different formation lengths and traction capacities in view of detailed passenger travel demands in different operation times or seasons of a rail transport system.

## 6. System Operation

The sustainable as well as efficient operation of a rail transport system is supported by not only superior performance of its equipment and facilities but also rational system operation management. However, because of, for example, spatially uneven transport demands as a whole, only aggregate statistics on the overall TEC intensity is unable to reflect the demand variety at different parts of a rail transport system [62] and certainly hard to provide adequately effective energy-saving approaches in systematic concepts. Taking into account interactions, timetable adherences, and so forth of multitrains operating on the same rail line or network, research on minimizing their total TEC, especially in recent years, makes much effort to get solutions from the perspective of systematic operation, as explained in Figure 5. This kind of research is able to reasonably contribute to decreasing the overall TEC intensity of a rail transport system.

Kraay et al. [63] and Higgins et al. [64] have tried to optimize the train operation diagram by utilizing mathematical programming to adjust the speeds and priorities of trains in an integrated manner for reducing their delays and total TEC. In view of the transport characteristics of the trains operating on the railway network of China in reality, Peng et al. [65] establish a schedule optimization model with the multiobjectives of optimal or near optimum total TTE and TEC of trains. The practicability and adaptability of the proposed model for complex network structure are guaranteed, and a decomposition algorithm is put forward to solve this model. In consideration of both structure complexity of a rail network and difference of passenger carrying capacities of different types of trains, Ghoseiri et al. [66] propose a new train scheduling model with more objectives besides decreasing the total TEC and TTE of trains, and the Pareto optimality [67] is suggested to be used to solve the model.

To reduce the TEC of a rail transport system and meanwhile maintain a certain service quality, dwell times of trains at stations or/and their interstop TTEs need dynamic

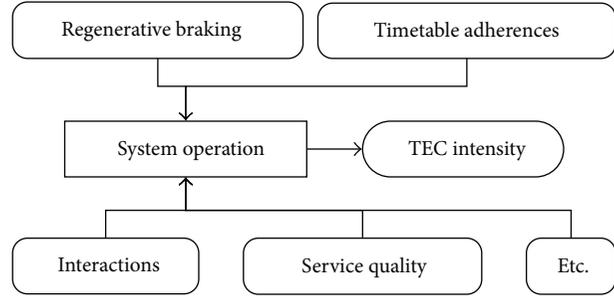


FIGURE 5: Rationally systematic operation for TEC saving.

adjustments to match time-varying travel demand. The dwell time adjustment is commonly utilized owing to its simplicity in practice [68]. For instance, a Heuristic-based train operation controller has been developed by Sansò and Girard [69] to reduce the power demand of a metro system in peak hours by introducing optimal dwell times for trains at successive stations. Because adjusting the interstop TTEs of a train with coasting controls can easily realize the exchange between its TEC and TTE for a whole trip [70], Wong and Ho [68] apply DP based on an event-based model to rationally optimize the interstop TTEs of metro trains according to dynamically changed passenger travel demand over a regional level especially in off-peak hours for energy saving under the premise of ensuring a certain level of the transport service.

If trains running on different rail lines are approaching the same transport section, the green wave strategy [71] for reducing their brakes and stops before entering the conflicting section is able to effectively decrease the TECs of these trains for their transports [72]. In order to make trains run through a conflicting transport section in the smoothest way, Bai et al. [73] apply golden section search method [74] to determine the most suitable speeds of trains when they are approaching a conflicting section. In comparison with the maximum traction strategy with the principle of improving the speed of a train before running into a conflicting transport section to its target speeds as much as possible, the proposed approach may save about 6% of the TEC of such a train. Moreover, Fu et al. [75] make use of a GA with changeable lengths of chromosomes to analyze the interimpacts of different trains passing through the same stop from different directions at nearly the same time. According to the analyzed inter-impacts, corresponding energy-saving train control rules are suggested.

In an integrated electricity supply viewpoint, Falvo et al. [76] study the energy-saving effect of regenerative braking [77] of metro trains running on a rail network. The utilizations of the electricity provided by the traction electric power supply system and the electric energy from the regenerative braking of metro trains for the tractions of trains on the same metro network are recorded in the simulation study of the transport operations of these trains. Regenerative braking enables much energy saving by not only injecting part of the braking kinetic energy of a train into the electrical grid to be consumed by nearby powering trains or returned to the AC

transmission system through a reversible substation [78] but also preventing the tunnel temperature rising in underground railways [79] to minimize the energy consumption in air conditioning or ventilating. The average utilization rate of the regenerative electric power for train tractions decreases with increasing the headways of multitrains even if all of them are equipped with electrical energy storage devices such as the flywheel system [80], and a train must activate the rheostatic braking to burn some regenerated power when the catenary voltage has reached its top limit [78]. The unused regenerative electric power for train traction is suggested to serve other electricity consumption systems, for example, electric vehicle charging stations in urban area, especially when the headways of the trains are relatively big.

In consideration of the interactions of trains in their following transports on the same rail line, Ding et al. [81] establish a dynamics model to optimize the energy-saving driving controls of following trains influenced by the dynamically changes of rail transport signals, according to their previously proposed simulation algorithm [82] for the energy-saving control of a single train under its fixed travel time constraint for a certain transport. The corresponding Heuristic algorithm is also proposed to estimate the dynamics model. Based on this model, the computer-aided simulation system is developed to get the optimal TEC of train following operation under the automatic block system [16]. This dynamics modeling research is capable of optimizing the controls of following trains according to the signals to avoid unnecessary brakes for adjusting their speeds. As a result, the TEC of such trains is decreased. Additionally, taking into account the energy saving effect of regenerative braking, Acikbas and Soylemez [83] make use of artificial neural networks [84, 85] to develop the simulation system taking a GA to search for systematically rational coasting start points of different trains. In a different manner, Yang et al. [86] first establish the penalty function to evaluate the influence of the interactions of trains upon their TTEs and TECs. Afterwards, the weight values of the overall TTE and TEC for an integrated cost evaluation are obtained by spreadsheet analyses. At last, a GA is accordingly utilized to explore the optimal control program in a systematic way.

Because of the uncertain delays of trains due to the inter-influence of their transports on the same rail line or network, Ding et al. [87] and Cucala et al. [88] combine coasting program optimization and transport schedule modeling [89, 90] to minimize the TEC of multitrains. In the transport simulation study, a GA is applied to explore energy-saving coasting schemes of the trains and obtain the relationship between the corresponding TEC and TTE of each scheme. In order to realize the least TEC for the trains resuming a transport time schedule in case of an uncertain delay, Ding et al. [87] distribute the slack travel time of the rail line to each of its transport sections simply according to the previously obtained relationship between the TEC and the TTE mainly in view of the track alignments in all the transport sections and the length of every stopspacing. In contrast, based on the relationship between the TEC and the TTE, the fuzzy linear programming [91] is utilized by Cucala et al. [88] to rationally make use of the slack time of a rail line for the delays which

are modeled as fuzzy numbers and punctuality constraints and capable of getting reasonable behavioral responses of the drivers for the most effective decrease of the general TEC. Such an approach is also applied to some Spanish HSR lines to evaluate the energy-saving potential of its commercial transport service.

In view of the TECs and carbon emissions of trains and the TTEs of their passengers, Li et al. [92] recently develop a green train scheduling model to optimize the transport timetable of multitrains in the operation of a rail transport system for the reduction of the system cost. The TEC of each train is determined according to the correlations between its running resistance, speed, and TTE for a transport work. The carbon emission cost of all the trains is computed based on their allowed and actual emission volumes. The total TTE of the passengers is used to describe the travel demand in a certain time period. This multiobjective nonlinear optimization problem is solved with the utilization of fuzzy mathematical programming [91]. It is proved that taking into account the TTE of all the passengers of multitrains operating on the same rail line or network, the newly developed model is able to effectively decrease the TEC and carbon emission of the systematic transport operation.

## 7. Summary

There is no doubt that as to the driving strategy optimization on the microlevel, a rational control program, especially a valid coasting scheme, of a train enables an evident decrease of its TEC for the same transport work. As a result, many studies have been made to explore reasonable approaches that are able to efficiently determine in particular the optimal coasting start points of a train in detail for its certain trip to ultimately realize dynamic optimization of its driving tactics on board for energy saving. Meanwhile, impacts of various types of track alignments on the TEC intensity of a train need adequate analyses in its operating control optimization. Moreover, it is also clarified on the mesolevel that the ameliorative attributes such as reduced weight, improved power output efficiency, decreased accelerating-braking frequency, and so forth of a train all play important roles in decreasing the TEC of the whole transport operation process of the train. Furthermore, from the perspective of the transport organization on the macrolevel, the interinfluence of multitrains in their coordinate operations requires systematic solutions to radically decrease the overall TEC intensity of a rail transport system in an effective manner.

In view of both instability of a system when its behavior influenced by multifactors changes seriously [93] and the systematic characteristics of the transport operation of a rail line or network, dynamically systematical optimization of the control programs of trains with reasonably amendatory attributes in sound responses to track alignments is necessary to be further studied in an integrated way as much as possible in future research on traction energy saving of rail transport. This is also much beneficial to the improvement of rail transport safety [94, 95]. Of course, utilizing new energy, for

example, hydrogen against electrification [96, 97], deserves our continual effort as well.

## Acknowledgments

This research is supported by the National Natural Science Foundation of China (71201006 and 71131001), the National Basic Research Program of China (2012CB725406) and the Fundamental Research Funds for the Central Universities (2011JBM251).

## References

- [1] P. C. M. Leung and E. W. M. Lee, "Estimation of electrical power consumption in subway station design by intelligent approach," *Applied Energy*, vol. 101, pp. 634–643, 2013.
- [2] C. Winston and V. Maheshri, "On the social desirability of urban rail transit systems," *Journal of Urban Economics*, vol. 62, no. 2, pp. 362–382, 2007.
- [3] J. Hickman, D. Hassel, R. Joumard, Z. Samaras, and S. Sorenson, *Methodology for Calculating Transport Emissions and Energy Consumption*, Office for Official Publications of the European Communities, Luxembourg, Belgium, 1999.
- [4] X. Feng, "Energy cost of high-speed railway trains in the effect of utilization ratios of passenger seats," in *Proceedings of the International Conference on Information, Services and Management Engineering (ISME '11)*, vol. 1, pp. 133–136, Beijing, China, December 2011.
- [5] E. Lindgreen and S. C. Sorenson, *Simulation of Energy Consumption and Emissions from Rail Traffic Evaluation*, Technical University of Denmark, Lyngby, Denmark, 2005.
- [6] P. Lukaszewicz, *Energy consumption and running time for trains: modelling of running resistance and driver behavior based on full scale testing [Ph.D. thesis]*, Royal Institute of Technology, Stockholm, Sweden, 2001.
- [7] A. R. Miller, J. Peters, B. E. Smith, and O. A. Velev, "Analysis of fuel cell hybrid locomotives," *Journal of Power Sources*, vol. 157, no. 2, pp. 855–861, 2006.
- [8] H. Wang, H. Hao, X. Li, K. Zhang, and M. Ouyang, "Performance of Euro III common rail heavy duty diesel engine fueled with gas to liquid," *Applied Energy*, vol. 86, no. 10, pp. 2257–2261, 2009.
- [9] E. V. Hoyt and R. R. Levary, "Assessing the effects of several variables on freight train fuel consumption and performance using a train performance simulator," *Transportation Research A*, vol. 24, no. 2, pp. 99–112, 1990.
- [10] R. S. Raghunathan, H. D. Kim, and T. Setoguchi, "Aerodynamics of high-speed railway train," *Progress in Aerospace Sciences*, vol. 38, no. 6-7, pp. 469–514, 2002.
- [11] H. I. Andrews, *Railway Traction: The Principles of Mechanical and Electrical Railway Traction*, vol. 5 of *Studies in Mechanical Engineering*, Elsevier Scientific, New York, NY, USA, 1986.
- [12] C. Chandra and M. M. Aqarwal, *Railway Engineering*, Oxford Higher Education, New York, NY, USA, 2008.
- [13] A. Chui, K. K. Li, and P. K. Lau, "Traction energy management in KCR," in *Proceedings of the 2nd International Conference on Advances in Power System Control, Operation & Management*, vol. 1, pp. 202–208, December 1993.
- [14] C. S. Chang and S. S. Sim, "Optimising train movements through coast control using genetic algorithms," *IEE Proceedings—Electric Power Application*, vol. 144, no. 1, pp. 65–73, 1997.
- [15] J. S. Arora, "Genetic algorithms for optimum design," in *Introduction to Optimum Design*, J. S. Arora, Ed., pp. 643–655, Academic Press, Waltham, UK, 3rd edition, 2012.
- [16] B. Solomon, *Railroad Signaling*, Voyageur Press, Minneapolis, Minn, USA, 2010.
- [17] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, Mass, USA, 3rd edition, 2009.
- [18] K. K. Wong and T. K. Ho, "Coast control for mass rapid transit railways with searching methods," *IEE Proceedings: Electric Power Applications*, vol. 151, no. 3, pp. 365–375, 2004.
- [19] H. Hwang, "Control strategy for optimal compromise between trip time and energy consumption in a high-speed railway," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 28, no. 6, pp. 791–802, 1998.
- [20] F. Hppner, F. Klawonn, K. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, John Wiley & Sons, Chichester, UK, 1999.
- [21] M. S. Yang, "A survey of fuzzy clustering," *Mathematical and Computer Modelling*, vol. 18, no. 11, pp. 1–16, 1993.
- [22] R. Liu and I. M. Golovitcher, "Energy-efficient operation of rail vehicles," *Transportation Research A*, vol. 37, no. 10, pp. 917–932, 2003.
- [23] Y. Bai, B. Mao, Y. Ding, F. Zhou, and W. Jia, "An onboard optimal control system for freight trains," in *Proceedings of the 6th International Conference on Traffic and Transportation Studies Congress*, pp. 770–783, Nanning, China, August 2008.
- [24] A. Ioffe and V. Tikhomirov, *Theory of External Problems*, Nauka, Moscow, Russia, 1974.
- [25] R. E. Showalter, "The maximum principle," in *Introduction to Holomorphy*, J. A. Barroso, Ed., vol. 106 of *North-Holland Mathematics Studies*, chapter 10, pp. 115–117, Elsevier Science B.V., Amsterdam, The Netherlands, 1985.
- [26] Y. V. Bocharnikov, A. M. Tobias, C. Roberts, S. Hillmansen, and C. J. Goodman, "Optimal driving strategy for traction energy saving on DC suburban railways," *IET Electric Power Applications*, vol. 1, no. 5, pp. 675–682, 2007.
- [27] T. R. Kane and D. A. Levinson, *Dynamics: Theory and Applications*, McGraw-Hill, New York, NY, USA, 1985.
- [28] J. N. Mordeson, "Fuzzy mathematics," in *Foundations of Image Understanding*, L. S. Davis, Ed., pp. 95–126, Springer Science, Business Media, New York, NY, USA, 2001.
- [29] K. Kim and S. I. Chien, "Simulation-based analysis of train controls under various track alignments," *Journal of Transportation Engineering*, vol. 136, no. 11, pp. 937–948, 2010.
- [30] D. Bertsimas and J. Tsitsiklis, "Simulated annealing," *Statistical Science*, vol. 8, no. 1, pp. 10–15, 1993.
- [31] K. Kim and S. I. Chien, "Optimal train operation for minimum energy consumption considering track alignment, speed limit, and schedule adherence," *Journal of Transportation Engineering*, vol. 137, no. 9, pp. 665–674, 2011.
- [32] G. F. Ignacio and G. A. Alberto, "Can high-speed trains run faster and reduce energy consumption," *Procedia—Social and Behavioral Sciences*, vol. 48, pp. 827–837, 2012.
- [33] P. Howlett, "The optimal control of a train," *Annals of Operations Research*, vol. 98, no. 1–4, pp. 65–87, 2000.

- [34] P. G. Howlett, P. J. Pudney, and X. Vu, "Local energy minimization in optimal train control," *Automatica*, vol. 45, no. 11, pp. 2692–2698, 2009.
- [35] H. Liu, B. Mao, Y. Ding, W. Jia, and S. Lai, "Train energy-saving scheme with evaluation in urban mass transit systems," *Journal of Transportation Systems Engineering and Information Technology*, vol. 7, no. 5, pp. 68–73, 2007.
- [36] H. H. Hoang, M. P. Polis, and A. Haurie, "Reducing energy consumption through trajectory optimization for a metro network," *IEEE Transactions on Automatic Control*, vol. 20, no. 5, pp. 590–595, 1975.
- [37] S. Edelkamp and S. Schrdl, *Heuristic Search: Theory and Applications*, Morgon Kaufmann, Waltham, Mass, USA, 2011.
- [38] E. V. Denardo, *Dynamic Programming: Models and Applications*, Dover, New York, NY, USA, 2003.
- [39] P. Firpo and S. Savio, "Optimized train running curve for electrical energy saving in autotransformer supplied AC railway systems," in *Proceedings of the International Conference on Electric Railways in a United Europe*, pp. 23–27, Amsterdam, The Netherlands, March 1995.
- [40] S. E. Lyshevski, *Electromechanical Systems, Electric Machines, and Applied Mechatronics*, CRC Press, Boca Raton, Fla, USA, 1999.
- [41] I. Maros, *Computational Techniques of the Simplex Method*, Kluwer Academic, Dordrecht, The Netherlands, 2002.
- [42] D. A. Coley, *An Introduction to Genetic Algorithms for Scientists and Engineers*, World Scientific, Singapore, 1999.
- [43] A. W. Johnson and S. H. Jacobson, "On the convergence of generalized hill climbing algorithms," *Discrete Applied Mathematics*, vol. 119, no. 1-2, pp. 37–57, 2002.
- [44] D. N. Kim and P. M. Schonfeld, "Benefits of dipped vertical alignments for rail transit routes," *Journal of Transportation Engineering*, vol. 123, no. 1, pp. 20–27, 1997.
- [45] B. R. Ke, M. C. Chen, and C. L. Lin, "Block-layout design using maxmin ant system for saving energy on mass rapid transit systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 226–235, 2009.
- [46] T. Stützle and H. H. Hoos, "MAX-MIN ant system," *Future Generation Computer Systems*, vol. 16, no. 8, pp. 889–914, 2000.
- [47] B. R. Ke, C. L. Lin, and C. W. Lai, "Optimization of train-speed trajectory and control for mass rapid transit systems," *Control Engineering Practice*, vol. 19, no. 7, pp. 675–687, 2011.
- [48] B. Liu, "Fuzzy process, hybrid process and uncertain process," *Journal of Uncertain Systems*, vol. 2, no. 1, pp. 3–16, 2008.
- [49] K. M. Passino and S. Yurkovich, *Fuzzy Control*, Addison-Wesley, Longman, Menlo Park, Calif, USA, 1998.
- [50] K. P. Kokotovic and S. G. Singh, "Minimum-energy control of a traction motor," *IEEE Transactions on Automatic Control*, vol. 17, no. 1, pp. 92–95, 1972.
- [51] A. Blokhuis, "Hamiltonian systems," in *Methods of Differential Geometry in Analytical Mechanics*, M. Len and P. R. Rodrigues, Eds., vol. 158 of *North-Holland Mathematics Studies*, chapter 6, pp. 263–299, Elsevier Science B.V., Amsterdam, The Netherlands, 1989.
- [52] R. H. Martin and M. Pierre, "The adjoint operator method," in *Computational Methods in Engineering Boundary Value Problems*, T. Y. Na, Ed., vol. 145 of *Mathematics in Science and Engineering*, chapter 4, pp. 52–69, Academic Press, New York, NY, USA, 1979.
- [53] A. Miele, "Extremization of linear integrals by green's theorem," in *Optimization Techniques with Applications to Aerospace Systems*, G. Leitmann, Ed., vol. 5 of *Mathematics in Science and Engineering*, pp. 69–98, Academic Press, New York, NY, USA, 1962.
- [54] IFEU (Institut für Energieund Umweltforschung Heidelberg GmbH), *Energy Savings by Light-Weighting (Final Report)*, International Aluminium Institute (IAI), Heidelberg, Germany, 2003.
- [55] M. Chou and X. Xia, "Optimal cruise control of heavy-haul trains equipped with electronically controlled pneumatic brake systems," *Control Engineering Practice*, vol. 15, no. 5, pp. 511–519, 2007.
- [56] S. Lu, S. Hillmansen, and C. Roberts, "A power-management strategy for multiple-unit railroad vehicles," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 2, pp. 406–420, 2011.
- [57] Y. Zhu, Y. Chen, G. Tian, H. Wu, and Q. Chen, "A four-step method to design an energy management strategy for hybrid vehicles," in *Proceedings of the American Control Conference (AAC '04)*, pp. 156–161, Boston, Mass, USA, July 2004.
- [58] X. Feng, "Optimization of target speeds of high-speed railway trains for traction energy saving and transport efficiency improvement," *Energy Policy*, vol. 39, no. 12, pp. 7658–7665, 2011.
- [59] X. Feng, B. Mao, X. Feng, and J. Feng, "Study on the maximum operation speeds of metro trains for energy saving as well as transport efficiency improvement," *Energy*, vol. 36, no. 11, pp. 6577–6582, 2011.
- [60] R. A. Uher, N. Sathi, and A. Sathi, "Traction energy cost reduction of the WMATA metro rail system," *IEEE Transactions on Industry Applications*, vol. IA-20, no. 3, pp. 472–483, 1984.
- [61] X. Feng, J. Feng, K. Wu, H. Liu, and Q. Sun, "Evaluating target speeds of passenger trains in China for energy saving in the effect of different formation scales and traction capacities," *International Journal of Electrical Power and Energy Systems*, vol. 42, no. 1, pp. 621–626, 2012.
- [62] I. López, J. Rodríguez, J. M. Burón, and A. García, "A methodology for evaluating environmental impacts of railway freight transportation policies," *Energy Policy*, vol. 37, no. 12, pp. 5393–5398, 2009.
- [63] D. Kraay, P. T. Harker, and B. Chen, "Optimal pacing of trains in freight railroads: model formulation and solution," *Operations Research*, vol. 39, no. 1, pp. 82–99, 1991.
- [64] A. Higgins, E. Kozan, and L. Ferreira, "Optimal scheduling of trains on a single line track," *Transportation Research B*, vol. 30, no. 2, pp. 147–161, 1996.
- [65] Q. Peng, S. Zhu, and P. Wang, "Study on the strategy of train operation adjustment on high speed railway," *Journal of the China Railway Society*, vol. 23, no. 1, pp. 1–8, 2001.
- [66] K. Ghoseiri, F. Szidarovszky, and M. J. Asgharpour, "A multi-objective train scheduling model and solution," *Transportation Research B*, vol. 38, no. 10, pp. 927–952, 2004.
- [67] M. Brio, G. M. Webb, and A. R. Zakharian, "Pareto optimality," in *The Vector-Valued Maximin*, V. I. Zhukovskiy and M. E. Salukvadze, Eds., vol. 193 of *Mathematics in Science and Engineering*, chapter 3, pp. 81–129, Academic Press, New York, NY, USA, 1994.
- [68] K. K. Wong and T. K. Ho, "Dwell-time and run-time control for DC mass rapid transit railways," *IET Electric Power Applications*, vol. 1, no. 6, pp. 956–966, 2007.

- [69] B. Sansò and P. Girard, "Train scheduling desynchronization and power peak optimization in a subway system," in *Proceedings of the IEEE/ASME Joint Railroad Conference*, pp. 75–78, Baltimore, Md, USA, April 1995.
- [70] K. K. Wong and T. K. Ho, "Dynamic coast control of train movement with genetic algorithm," *International Journal of Systems Science*, vol. 35, no. 13-14, pp. 835–846, 2004.
- [71] T. Albrecht, "The influence of anticipating train driving on the dispatching process in railway conflict situations," *Networks and Spatial Economics*, vol. 9, no. 1, pp. 85–101, 2009.
- [72] F. Corman, A. D'Ariano, D. Pacciarelli, and M. Pranzo, "Evaluation of green wave policy in real-time railway traffic management," *Transportation Research C*, vol. 17, no. 6, pp. 607–616, 2009.
- [73] Y. Bai, T. O. Ho, and B. Mao, "Train control to reduce delays upon service disturbances at railway junctions," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 5, pp. 114–122, 2011.
- [74] W. J. Braun and D. J. Murdoch, *A First Course in Statistical Programming with R*, Cambridge University Press, Cambridge, UK, 2007.
- [75] Y. Fu, Z. Gao, and K. Li, "Optimization method of energy saving train operation for railway network," *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 4, pp. 90–96, 2009.
- [76] M. C. Falvo, R. Lamedica, R. Bartoni, and G. Maranzano, "Energy management in metro-transit systems: an innovative proposal toward an integrated and sustainable urban mobility system including plug-in electric vehicles," *Electric Power Systems Research*, vol. 81, no. 12, pp. 2127–2138, 2011.
- [77] S. Robertson and J. Markham, *Regenerative Braking Story*, Venture Publications, Glossop, UK, 2007.
- [78] M. Coto, P. Arbolea, and C. Gonzalez-Moran, "Optimization approach to unified AC/DC power flow applied to traction systems with catenary voltage constraints," *International Journal of Electrical Power and Energy Systems*, vol. 53, no. 1, pp. 434–441, 2013.
- [79] T. Suzuki, "DC power-supply system with inverting substations for traction systems using regenerative brakes," *IEEE Proceedings B*, vol. 129, no. 1, pp. 18–26, 1982.
- [80] M. B. Richardson, "Flywheel energy storage system for traction applications," in *Proceedings of the International Conference on Power Electronics, Machines and Drives*, Conference Publication no. 487, pp. 275–279, Bath, UK, April 2002.
- [81] Y. Ding, F. Zhou, Y. Bai, and H. Liu, "Simulation algorithm for energy-efficient train following operation under automatic-block system," *Journal of System Simulation*, vol. 21, no. 15, pp. 4593–4597, 2009.
- [82] Y. Ding, B. Mao, H. Liu, T. Wang, and X. Zhang, "Algorithm for energy-efficient train operation simulation with fixed running time," *Journal of System Simulation*, vol. 16, no. 10, pp. 2241–2244, 2004.
- [83] S. Acikbas and M. T. Soylemez, "Coasting point optimisation for mass rail transit lines using artificial neural networks and genetic algorithms," *IET Electric Power Applications*, vol. 2, no. 3, pp. 172–182, 2008.
- [84] D. Anderson and G. McNeill, *Artificial Neural Networks Technology: A DACS State-of-the-Art Report*, Kaman Sciences Corporation, New York, NY, USA, 1992.
- [85] D. J. Livingstone, *Artificial Neural Networks: Methods and Applications*, Humana Press, Totowa, NJ, USA, 2008.
- [86] L. Yang, K. Li, Z. Gao, and X. Li, "Optimizing trains movement on a railway network," *Omega*, vol. 40, no. 5, pp. 619–633, 2012.
- [87] Y. Ding, H. Liu, Y. Bai, and F. Zhou, "A two-level optimization model and algorithm for energy-efficient urban train operation," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 1, pp. 96–101, 2011.
- [88] A. P. Cucala, A. Fernández, C. Sicre, and M. Domínguez, "Fuzzy optimal schedule of high speed train operation to minimize energy consumption with uncertain delays and driver's behavioral response," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 8, pp. 1548–1557, 2012.
- [89] M. Hickman, P. Mirchandani, and S. Vo, *Computer-Aided Systems in Public Transport*, Springer, Berlin, Germany, 2008.
- [90] M. Salicrú, C. Fleurent, and J. M. Armengol, "Timetable-based operation in urban transport: run-time optimisation and improvements in the operating process," *Transportation Research A*, vol. 45, no. 8, pp. 721–740, 2011.
- [91] Y. J. Lai and C. L. Hwang, *Fuzzy Mathematical Programming: Methods and Applications*, Springer, New York, NY, USA, 1995.
- [92] X. Li, D. Wang, K. Li, and Z. Gao, "A green train scheduling model and fuzzy multi-objective optimization algorithm," *Applied Mathematical Modelling*, vol. 37, no. 4, pp. 2063–2073, 2013.
- [93] W. Wang, X. Jiang, S. Xia, and Q. Cao, "Incident tree model and incident tree analysis method for quantified risk assessment: an in-depth accident study in traffic operation," *Safety Science*, vol. 48, no. 10, pp. 1248–1262, 2010.
- [94] M. Nejlaoui, A. Houidi, Z. Affi, and L. Romdhane, "Multiobjective robust design optimization of rail vehicle moving in short radius curved tracks based on the safety and comfort criteria," *Simulation Modelling Practice and Theory*, vol. 30, pp. 21–34, 2013.
- [95] E. Matsika, S. Ricci, P. Mortimer, N. Georgiev, and C. O'Neill, "Rail vehicles, environment, safety and security," *Research in Transportation Economics*, vol. 41, no. 1, pp. 43–58, 2013.
- [96] G. D. Marin, G. F. Naterer, and K. Gabriel, "Rail transportation by hydrogen versus electrification—case study for Ontario Canada—I: propulsion and storage," *International Journal of Hydrogen Energy*, vol. 35, no. 12, pp. 6084–6096, 2010.
- [97] G. D. Marin, G. F. Naterer, and K. Gabriel, "Rail transportation by hydrogen versus electrification—case study for Ontario, Canada—II: energy supply and distribution," *International Journal of Hydrogen Energy*, vol. 35, no. 12, pp. 6097–6107, 2010.