

Discrete Dynamics in Nature and Society

An International Multidisciplinary Research and Review Journal

Special Issue
Discrete Dynamics in Transportation System

Guest Editors: Wuhong Wang, Klaus Bengler, and Geert Wets



Discrete Dynamics in Transportation System

Discrete Dynamics in Nature and Society

Discrete Dynamics in Transportation System

Guest Editors: Wuhong Wang, Klaus Bengler,
and Geert Wets



Copyright © 2013 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 Ait 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
M. 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 F.-Prunaret, France
Vladimir Gontar, Israel
Taher S. Hassan, Egypt
Xue Zhong He, Australia
Weihong Huang, Singapore
Giuseppe Izzo, Italy
S. Jagannathan, USA
Lee Chae Jang, Korea
Jun Ji, USA
Zhen Jin, China
N. I. Karachalios, Greece
Eric R. Kaufmann, USA
Recai Kilic, Turkey
- Ryusuke Kon, Japan
Risto Korhonen, Finland
Victor S. Kozyakin, Russia
M. R. S. Kulenovic, USA
J. Kurths, Germany
WeiNian 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 Huu Anh Ngoc, Vietnam
Piyapong Niamsup, Thailand
Juan J. Nieto, Spain
Prasanta K. Panigrahi, India
G. Papaschinopoulos, Greece
Beatrice Paternoster, Italy
Mingshu Peng, China
Carlo Piccardi, Italy
Peter Plath, Germany
Chuanxi Qian, USA
V. S. Rabinovich, Mexico
I. Rachunková, Czech Republic
M. Rafei, The Netherlands
Aura Reggiani, Italy
Pavel Rehak, Czech Republic
Marko Robnik, Slovenia
Yuriy Rogovchenko, Norway
Ventsi G. Rumchev, Australia
B. S. Daya Sagar, India
- R. Sahadevan, India
Leonid Shaikhnet, Ukraine
Vimal Singh, Turkey
Seenith Sivasundaram, USA
Francisco Solis, Mexico
G. Solís-Perales, Mexico
Yihong Song, China
M. Sonis, Israel
Jian-Ping Sun, China
Stepan Agop Tersian, Bulgaria
Gerald Teschl, Austria
Tetsuji Tokihiro, Japan
Delfim F. M. Torres, Portugal
Firdaus Udwardia, USA
Antonia Vecchio, Italy
Qiru Wang, China
Ibrahim Yalcinkaya, Turkey
Xiang Ping Yan, China
Bo Yang, USA
Xiaofan Yang, China
Her-Terng Yau, Taiwan
Jianshe Yu, China
Ağacık Zafer, Turkey
Binggen Zhang, China
Guang Zhang, China
Zhengqiu Zhang, China
Lu Zhen, China
Baodong Zheng, China
Yong Zhou, China
Zhan Zhou, China
Zuo-nong Zhu, China

Contents

Discrete Dynamics in Transportation System, Wuhong Wang, Klaus Bengler, and Geert Wets
Volume 2013, Article ID 234970, 2 pages

Impacts of Snowy Weather Conditions on Expressway Traffic Flow Characteristics,
Jiancheng Weng, Lili Liu, and Jian Rong
Volume 2013, Article ID 791743, 6 pages

Optimal Energy Control Strategy Design for a Hybrid Electric Vehicle, Yuan Zou, Hou Shi-jie,
Li Dong-ge, Gao Wei, and Xiao-song Hu
Volume 2013, Article ID 132064, 8 pages

**A Pilot Study Verifying How the Curve Information Impacts on the Driver Performance with
Cognition Model**, Xiaohua Zhao, Wei Guan, and Xiaoming Liu
Volume 2013, Article ID 316896, 8 pages

Modeling Lane-Keeping Behavior of Bicyclists Using Survival Analysis Approach,
Hongwei Guo, Wuhong Wang, Weiwei Guo, and Facheng Zhao
Volume 2013, Article ID 197518, 6 pages

Individual Subjective Initiative Merge Model Based on Cellular Automaton, Yin-Jie Xu,
Yu-Guang Chen, Chao Yang, and Yi-Chao Pu
Volume 2013, Article ID 320943, 7 pages

**Emissions and Fuel Consumption Modeling for Evaluating Environmental Effectiveness of
ITS Strategies**, Yuan-yuan Song, En-jian Yao, Ting Zuo, and Zhi-feng Lang
Volume 2013, Article ID 581945, 9 pages

A Real-Time Lane Detection Algorithm Based on Intelligent CCD Parameters Regulation,
Ping-shu Ge, Lie Guo, Guo-kai Xu, Rong-hui Zhang, and Tao Zhang
Volume 2012, Article ID 273164, 16 pages

A Cloud-Computing-Based Data Placement Strategy in High-Speed Railway, Hanning Wang,
Weixiang Xu, Futian Wang, and Chaolong Jia
Volume 2012, Article ID 396387, 15 pages

Comprehensive Evaluation of Driver's Propensity Based on Evidence Theory, Jinglei Zhang,
Xiaoyuan Wang, Song Gao, and Yiqing Song
Volume 2012, Article ID 103582, 8 pages

Concept Layout Model of Transportation Terminals, Li-ya Yao, Li-shan Sun, Wu-hong Wang,
and Hui Xiong
Volume 2012, Article ID 148216, 8 pages

Offset Optimization Based on Queue Length Constraint for Saturated Arterial Intersections,
Xianmin Song, Pengfei Tao, Ligang Chen, and Dianhai Wang
Volume 2012, Article ID 907639, 13 pages

**An Improved Model for Headway-Based Bus Service Unreliability Prevention with Vehicle
Load Capacity Constraint at Bus Stops**, Weiya Chen, Chunhua Yang, Fenling Feng,
and Zhiya Chen
Volume 2012, Article ID 313518, 13 pages

Forecasting Cohesionless Soil Highway Slope Displacement Using Modular Neural Network, Yanyan Chen, Shuwei Wang, Ning Chen, Xueqin Long, and Xiru Tang
Volume 2012, Article ID 504574, 15 pages

A Fuzzy Optimization Model for High-Speed Railway Timetable Rescheduling, Li Wang, Yong Qin, Jie Xu, and Limin Jia
Volume 2012, Article ID 827073, 22 pages

Lane-Level Vehicle Trajectory Reckoning for Cooperative Vehicle-Infrastructure System, Yinsong Wang, Xiaoguang Yang, Luoyi Huang, and Jiawen Wang
Volume 2012, Article ID 941047, 15 pages

Analysis of ITMS System Impact Mechanism in Beijing Based on FD and Traffic Entropy, Ailing Huang, Wei Guan, Yimei Chang, and Zhen Yang
Volume 2012, Article ID 350847, 16 pages

Fast Pedestrian Recognition Based on Multisensor Fusion, Hongyu Hu, Zhaowei Qu, Zhihui Li, Jinhui Hu, and Fulu Wei
Volume 2012, Article ID 318305, 14 pages

A Novel Short-Range Prediction Model for Railway Track Irregularity, Peng Xu, Rengkui Liu, Quanxin Sun, and Futian Wang
Volume 2012, Article ID 591490, 12 pages

Study on Optimization of Night Illumination in Expressway Long Tunnels, Weihua Zhao, Hao Chen, Qiang Yu, and Haoxue Liu
Volume 2012, Article ID 160436, 9 pages

Pedestrian Walking Behavior Revealed through a Random Walk Model, Hui Xiong, Liya Yao, Huachun Tan, and Wuhong Wang
Volume 2012, Article ID 405907, 13 pages

Analysis of the Travel Intent for Park and Ride Based on Perception, Huanmei Qin, Hongzhi Guan, and Guang Zhang
Volume 2012, Article ID 516197, 14 pages

A Numerical Model for Railroad Freight Car-to-Car End Impact, Chao Chen, Mei Han, and Yanhui Han
Volume 2012, Article ID 927592, 11 pages

Traffic Behavior in CA Model of Vehicular Traffic through a Series of Signals, Kazuhiro Tobita and Takashi Nagatani
Volume 2012, Article ID 812969, 17 pages

Track Irregularity Time Series Analysis and Trend Forecasting, Jia Chaolong, Xu Weixiang, Wang Futian, and Wang Hanning
Volume 2012, Article ID 387857, 15 pages

Research on Analysis Method of Traffic Congestion Mechanism Based on Improved Cell Transmission Model, Hongzhao Dong, Shuai Ma, Mingfei Guo, and Dongxu Liu
Volume 2012, Article ID 854654, 11 pages

Dynamic Recognition Model of Drivers Propensity under Multilane Traffic Environments, Xiaoyuan Wang, Jin Liu, and Jinglei Zhang
Volume 2012, Article ID 309415, 15 pages

Distinguishing between Rural and Urban Road Segment Traffic Safety Based on Zero-Inflated Negative Binomial Regression Models, Xuedong Yan, Bin Wang, Meiwu An, and Cuiping Zhang
Volume 2012, Article ID 789140, 11 pages

Driver Cognitive Distraction Detection Using Driving Performance Measures, Lisheng Jin, Qingning Niu, Haijing Hou, Huacai Xian, Yali Wang, and Dongdong Shi
Volume 2012, Article ID 432634, 12 pages

Transportation Structure Analysis Using SD-MOP in World Modern Garden City: A Case Study in China, Jiuping Xu, Jing Yang, and Liming Yao
Volume 2012, Article ID 710854, 23 pages

A Signal Coordination Control Based on Traversing Empty between Mid-Block Street Crossing and Intersection, Changjiang Zheng, Genghua Ma, Jinhua Wu, Xiaoli Zhang, and Xiaolei Zhang
Volume 2012, Article ID 189316, 18 pages

Analysis on the Synergy Evolutionary Development of the Collecting, Distributing, and Transporting System of Railway Heavy Haul Transportation, Fenling Feng, Dan Lan, and Liuwen Yang
Volume 2012, Article ID 329120, 22 pages

A Study on the Coordination of Urban Traffic Control and Traffic Assignment, ZhaoWei Qu, Yan Xing, XianMin Song, YuZhou Duan, and Fulu Wei
Volume 2012, Article ID 367468, 12 pages

Estimation of Saturation Flow Rates at Signalized Intersections, Chang-qiao Shao and Xiao-ming Liu
Volume 2012, Article ID 720474, 9 pages

An Optimization to Schedule Train Operations with Phase-Regular Framework for Intercity Rail Lines, Huimin Niu and Minghui Zhang
Volume 2012, Article ID 549374, 13 pages

Coherent Network Optimizing of Rail-Based Urban Mass Transit, Ying Zhang
Volume 2012, Article ID 960762, 8 pages

Study on the Train Operation Optimization of Passenger Dedicated Lines Based on Satisfaction, Zhipeng Huang and Huimin Niu
Volume 2012, Article ID 451201, 11 pages

Stochastic User Equilibrium Assignment in Schedule-Based Transit Networks with Capacity Constraints, Wangtu Xu, Lixin Miao, and Wei-Hua Lin
Volume 2012, Article ID 910754, 15 pages

Crossing at a Red Light: Behavior of Cyclists at Urban Intersections, Xiaobao Yang, Mei Huan, Bingfeng Si, Liang Gao, and Hongwei Guo
Volume 2012, Article ID 490810, 12 pages

Successive-Stage Speed Limit on Exit Ramp Upstream of Direct-Type Freeway in China,

Hongwei Li, Jian Lu, Yongfeng Ma, and Yuanlin Liu

Volume 2012, Article ID 329507, 16 pages

A Model for Bus Crew Scheduling Problem with Multiple Duty Types, Mingming Chen and

Huimin Niu

Volume 2012, Article ID 649213, 11 pages

Research on Public Transit Network Hierarchy Based on Residential Transit Trip Distance,

Gao Jian, Zhao Peng, Zhuge Chengxiang, and Zhang Hui

Volume 2012, Article ID 390128, 14 pages

Evolving Model for the Complex Traffic and Transportation Network Considering Self-Growth Situation, Wei Zhang and Di Xu

Volume 2012, Article ID 291965, 12 pages

Handling Stability of Tractor Semitrailer Based on Handling Diagram, Ren Yuan-yuan,

Zheng Xue-lian, and Li Xian-sheng

Volume 2012, Article ID 350360, 16 pages

Editorial

Discrete Dynamics in Transportation System

Wuhong Wang,¹ Klaus Bengler,² and Geert Wets³

¹ Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

² Lehrstuhl für Ergonomie, Technische Universität München, 85747 Munich, Germany

³ Transportation Research Institute (IMOB), Hasselt University, 3590 Diepenbeek, Belgium

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

Received 24 October 2012; Accepted 31 October 2012

Copyright © 2013 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.

1. Overview

A transportation system is a dynamic, diverse, random and complex system, which consists of road users, vehicles, roadway, and the environment. An efficient transportation system is essential for the function and prosperity of modern, industrialized societies, and any disturbance of the components will transform the state of the whole system.

A transportation system exhibits extremely complex behavior derived from several components: the heterogeneous nature of human behavior, highly nonlinear group dynamics, and large system dimensions. Advances in mobility are most clearly illustrated by the spread of motorized traffic and transport and the development of the transportation industry, including vehicle manufacturing and its associated infrastructure. However, because of the growth of the transportation system, it also shows some negative impacts such as the increasing number of traffic accidents, the danger of health problems due to pollution, and the possibility of economical drawbacks due to congestion problems. Therefore, to reduce these negative impacts, the current developments in the transportation system represent one of the main challenges for information science and intelligent technology.

In order to reach a more efficient and safe road usage, some innovative techniques are performing in a transportation system, for example, intelligent transportation system (ITS). ITS includes a wide and growing suite of technologies and applications, aiming at providing available services relating to different modes of transportation and traffic management and enabling various users to make safer, more coordinated, and “smarter,” use of transportation system. The whole techniques of ITS can be divided into two parts:

stand-alone systems and cooperative systems. The two kinds of systems should work together to deliver the most efficient, safe, secure and comfortable journey.

To tackle the problems that were introduced before, traffic on an urban transportation system can be considered as a dynamical system. In recent years, some researches based on chaos theory were used to describe the traffic flow. Nonlinear chaotic phenomena was found in the short-term traffic flows, and even some implications were discussed for urban and transportation planning and forecasting. Period doubling, intermittency, and quasiperiodicity would make the transportation system into chaotic. The chaotic traffic flow model is a macroscopic model, which was established by adopting car following model. The model is a discrete and dynamic one derived from both the flow-density-speed fundamental diagram and the Greenshield's model, which use occupancy as variable and the ratio of free flow and average speed as control parameter.

With the development of society, the transport system shows a high complex and variational scene and enormous stochasticity. Various transportation problems and activities, such as accidents, unreasonable control signal, and temporary society actives, are derived from nonlinear singular factors which are the main causes of the stochasticity. This stochasticity can be defined as the impedance of the path. Consequently, the problem proposed how to find out the expected shortest path in a transport network, where the link travel times are modeled as a continuous-time stochastic process. The famous approaches for solving the problem are the Bellman's dynamic programming method for directed networks, the Dijkstra labeling method, and Bellman-Ford successive approximation method for networks with

nonnegative cost coefficients. Some important applications of the path optimization problem include vehicle routing in transportation systems, traffic routing in communication networks, and path planning in robotic systems.

Discrete optimization is a specific branch of optimization in applied mathematics and computer science. A very well-known application area of discrete optimization is a rail transport system. It helps us to plan and optimize the railroad network, in particular, the computation of the line plans, train schedules, and schedules of rolling stock. For example, the line planning consists in choosing a set of operating lines and its frequencies to serve the passenger demand and to optimize some given objective. However, the process of privatization of state-owned railroads enforces the efficient resource utilization. For this reason, cost-optimal line planning is modeled. Besides the routes and the frequencies, in the cost-optimal line planning approach, one has to find the optimal number of coaches per train, based on trivial estimates of the circulation of rolling stock.

To solve problems of a transportation system, next to optimization also human behavior plays a crucial role. Among various disturbances of transportation system, the reliability of human action in transportation system plays a crucial role in optimizing both person-task fit and safety issues. According to the result of all investigations over the world, accidents are to be led back to more than 95% exclusively on the driver's behavior. Disturbances such as human error can make the transportation system into a variety of unreliable state. On the basis of the measuring system disturbances, the unstable state can be defined as varying degrees of safety, such as conflict, near accident, or accident. Certainly, an unreliable transportation system can also transform into a reliable state by compensate strategies in safety efforts. The advanced driver assistance systems (ADASs) is such a subsystem of ITS, which helps drivers in perception, decision making, and action while driving and enhances the possibility of human error compensation, such as collision warning systems, lane departure warning systems, adaptive cruise control system (ACC), vision enhancement, and pedestrian detection.

The focus of this special issue is to foster links between basic and applied research relating to discrete dynamics problems in a transportation system. We invite original contributions on all topics in both new theoretical developments and studies of practical implementations, which propose and study novel dynamic and diverse models to reveal the mechanism of transportation system. The topics include intelligent transportation system (ITS), dynamic problem in urban transport system, transit and rail systems operation, application of chaos in traffic flow, discrete and stochastic theory of transport system, vehicle active safety and intelligent vehicle, discrete optimization methods in traffic system, traffic operations, management and control, energy, transport policy and economics, driving behavior, and driver assistance system.

2. Summary

An efficient and safe transportation system is essential for the function and prosperity of modern society. Nowadays, we have benefited from the rapid development of the

transportation industry. However, we also have to suffer from serious problems in transportation. Therefore, many researchers are seeking solutions to make a transportation system more efficient and improve the operations by using new technologies and new methodologies.

The analyzing, testing, modeling, and simulating of a transportation system and the state-of-the-art technologies will collectively improve the accidents, congestion, and pollution emissions problem of the transportation system.

This special issue aims at understanding the discrete state modeling the dynamic changes of the transportation system and finding a more reliable compensating approach to enhance the safety and efficiency of the transportation system. We believe that the papers in the special issue could be a representative study in the field of intelligent transportation system and help for a heuristic understanding of the current development in the field of transportation safety.

*Wuhong Wang
Klaus Bengler
Geert Wets*

Research Article

Impacts of Snowy Weather Conditions on Expressway Traffic Flow Characteristics

Jiancheng Weng, Lili Liu, and Jian Rong

Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, No. 100, Pingleyuan, Chaoyang District, Beijing 100124, China

Correspondence should be addressed to Jiancheng Weng; youthweng@bjut.edu.cn

Received 13 September 2012; Revised 11 January 2013; Accepted 11 January 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Jiancheng Weng 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.

Snowy weather will significantly degrade expressway operations, reduce service levels, and increase driving difficulty. Furthermore, the impact of snow varies in different types of roads, diverse cities, and snow densities due to different driving behavior. Traffic flow parameters are essential to decide what should be appropriate for weather-related traffic management and control strategies. This paper takes Beijing as a case study and analyzes traffic flow data collected by detectors in expressways. By comparing the performance of traffic flow under normal and snowy weather conditions, this paper quantitatively describes the impact of adverse weather on expressway volume and average speeds. Results indicate that average speeds on the Beijing expressway under heavy snow conditions decrease by 10–20 km/h when compared to those under normal weather conditions, the vehicle headway generally increases by 2–4 seconds, and the road capacity drops by about 33%. This paper also develops a specific expressway traffic parameter reduction model which proposes reduction coefficients of expressway volumes and speeds under various snow density conditions in Beijing. The conclusions paper provide effective foundational parameters for urban expressway controls and traffic management under snow conditions.

1. Introduction

The presence of fog, dust, rain, snow, or smoke in the atmosphere makes a driver's task more difficult. Since the middle of the twentieth century [1], it has been recognized that weather conditions have special influence on the road network operation due to the various behaviors of drivers under different weather conditions. Since the 1970s, several studies have focused on quantitative impact evaluations of the transportation system performance under adverse weather conditions in the USA [2–4]. They have confirmed that adverse weather conditions can significantly reduce the operating speed and thus the capacity in a given road segment [5]. However, few studies on road conditions and driver's behavior during inclement weather have been conducted in China. Due to the insufficient traffic performance analysis and traffic flow characteristic studies of traffic behavior in inclement weather in China, the implementation of traffic management measures, traffic control strategies, and traffic designs for inclement weather conditions has usually been

set and adjusted based only on a personal experience. This approach is quite impractical because it requires a traffic operator to manually change the settings in the field which deeply affects the reliability and efficiency of the traffic system. Specifically, the research on the impact of adverse weather in China has mainly focused on freeway management and traffic safety guarantees [6–9]. Adverse weather undoubtedly makes the driving operation more difficult, but it is still unclear how much and what changes should be made under various inclement weather conditions' even though a previous research based on empirical methods [10, 11] and simulation methods [12–14] has revealed that there is a substantial impact on driving in adverse weather.

Of all the adverse weather conditions, snow has the most significant effect on the driving behavior and vehicle performance because of the dual influence of road surfaces and visibility problems [15]. The snow scatters lights and creates a great amount of visual noise in a driver's visual field. In addition, if the snow is heavy enough, it covers the road surface, obscuring lane markings and road signs designed to

aid the driver. If the temperature is still low in the next day, the road surface freezes which then leads to a new threat to vehicles and traffic flow.

Several studies have examined the impact of snow and ice conditions specifically. McBride [2] evaluated the economic impact of highway controls and operation under snow and ice conditions and estimated the extra fuel consumption and the additional travel delay. Ibrahim and Hall [16] processed the traffic flow data collected from a freeway in Ontario, Canada, by using multivariable regression analysis methods and found that there was a 3%–5% speed reduction with light snow and a 30%–40% reduction in travel speeds with heavy snow. The Minnesota Department of Transportation also conducted a study based on data collected from a busy metropolitan freeway, highway 36, during rush hours between 3:00 and 8:00 PM on several weekdays under different weather conditions [11]. Their analysis showed that speeds dropped from 44 to 26 mph (40%), saturation flows decreased from 1800 to 1600 v/l/h (11%), and the start-up delays increased from 2 to 3 seconds.

Ren [17] also analyzed road sections and intersection run characteristics under a single snowy weather event and calculated the reduction percentages of road capacity, traffic volume, and travel speed at different intensity levels of snow. However, his conclusions were not universal and were inapplicable for other cities and other weather conditions.

Other authors have suggested potential solutions to help traffic flow during difficult weather conditions. Hu [14] discussed the impact of snowy weather on signal timing and evaluated the performance of intersection signals under icy and snowy weather conditions using a traffic simulation approach. Their simulation results indicated that a weather-responsive signal scheme can enhance traffic conditions, decrease travel delays, and raise travel speeds. Al-Kaisy and Freedman [18] presented a systematic investigation into the effects of adverse weather on signal operation and developed a set of recommended guidelines that relate weather conditions to operational impacts and potential benefits of weather-responsive signal timing.

The research also shows that drivers experience and behavior have an obvious influence on the traffic flow characteristics [19]. In most northern states in the USA, serious adverse weathers like snow storms occur frequently, so drivers are familiar with the conditions and have the experience driving under these adverse weathers, so they modify their driving habits accordingly. These modifications usually involve reducing speed and keeping greater distances between vehicles. Boyce's study illustrates that there are different influences and principles of snow's impact in different cities.

Some studies have also reported that the impact of weather conditions on traffic flow relationships and parameters is different depending on the road type. Chin et al. [20] established a traffic performance decrease table of various classes of roads under six different inclement weather conditions including light rain, heavy rain, light snow, heavy snow, fog, and ice. They used loop detector data gathered from different regions of the USA. From these data, it is clear

that not all weather conditions affect the traffic and driver's behavior equally.

In conclusion, a substantial amount of research has been conducted in other countries, providing models of significant reductive models in terms of capacity, travel, and signal timing; in contrast, studies on the impact of snowy weather in different cities and different road networks are inadequate and not systemic in China, and even less research has integrated traffic data supports. Therefore, few studies can respond to the questions of how to identify and evaluate traffic flow characteristics in snowy weather in Beijing and which countermeasures and traffic management strategies should be implemented such as access signal adjustments, expressways merge control, and lane closures.

The current study was initiated in response to this need. Specifically, this study attempted to analyze the traffic performance of expressways under different weather conditions in Beijing using the detector data. The results are revealed in three aspects including travel speed, traffic volume, and the saturation flow rate. Furthermore, a specific expressway traffic parameter reductive model under adverse weather conditions of various snow volume levels was developed. Finally, the main conclusions derived from the study are summarized.

2. Traffic Flow Data

The traffic flow data used in this study were collected from the microwave detectors which are installed on the expressways in Beijing. The expressway plays a significant role in the urban road network operation and accounts for more than 25% of the commuter travel mileages in Beijing. The time period of data collection was during the day (from 6:00 AM to 10:00 PM) on the snowy day with different snowfalls and the subsequent day. In order to enhance the data credibility, videotaped data were also collected over a period of several hours to calibrate the fixed detectors.

The time interval of data collection was 2 minutes, and the main fields of traffic flow data including travel speed, traffic volume, and occupancy were recorded in the database. The data were combined to analyze the traffic performance of road sections. The main fields of detector data are shown in Table 1.

A data before the process was implemented to eliminate data errors and faults and to improve the traffic data reliability since error data or lost data can result from mechanical errors of the detectors. The procedures including data filtering, data recovery, and data denoising are required to gain higher-quality data. Based on the processed data, the macroscopic indexes of traffic volume, travel speed, and density can be estimated and used to describe the basic characteristics of traffic flow under different weather conditions.

3. Speed Analyses

3.1. General Characteristics of Travel Speed under Snow Conditions. Traffic volume on expressways shows an obvious decline under heavy snowy weather conditions compared to

TABLE 1: Main fields of microwave detector data in Beijing.

Date	Time ID	Detector ID	Detected speed (km/h)	Occupancy	Traffic flow (veh.)
2009-12-27	655	HI7052c	33.6	2	14
2009-12-27	655	HI9141d	16.8	7	24
2009-12-27	655	HI9073d	39.8	3	21
2009-12-27	656	HI7052c	23.1	4	27

those under normal weather, and even the spot speed has a little increase [21]. To illustrate the general characteristics of traffic flow and speed under adverse weather conditions, the data was collected on January 3, 2010 with heavy snow (snow precipitation is greater than 5.0 mm) and December 27, 2009 with normal weather. The traffic flow data of each direction in twelve different points which located in the four orientations of the expressways including the second, the third, and the forth ring road in Beijing were analyzed. The observation time included multiple periods and was not restricted to the peak periods because of the unpredictable nature of snowfall intensity. The three-time periods were compared including the morning peak hours, the evening peak hours, and the off-peak hours.

The results are shown in Table 2. The average travel speed was the average value of the multiple groups' detected speed, and the section traffic volume was the average hourly observed volume during the specified time periods. From the result, it can be concluded that there was about 65% drop in traffic volume on the Beijing expressway under heavy snow conditions, but the travel speed was almost the same as the normal weather conditions due to the great drop of travel demand. The reason may be that there was a new balance between the decreased travel demand and deteriorated road supply condition.

Traffic volume and snow are the primary factors which affect the average velocity of the expressway on a snowy day because the road condition, geometric design, and time of day are all the same in different days. The heavy snowy day observed in this study was Sunday, so people had less rigid travel demands such as commuting. In addition, long-distance travel demand was likely canceled due to the heavy snow, and people only attempted to travel near their homes. Therefore, the traffic volume experienced a steep drop on the expressway from 65% to 72% because of the decreased long-distance travel. Thus, there was a large-scale drop in traffic volume, and mutual disturbances between vehicles were reduced, which even led to an increase in the peak hour and slight speed drop in off-peak hour on the identical road sections.

During off-peak hours, vehicles on the expressway also had high freedom to drive under the normal weather; consequently, the snow factor played a leading role in the speed change in snowy weather and caused a 10% speed reduction. In the evening peak hours, the traffic volume dropped over 70% resulting from the travel demand reduction. This alleviated the high-road pressure during the evening peak hours compared to normal weather conditions, which even enabled the average speed of the expressway to increase by 20%.

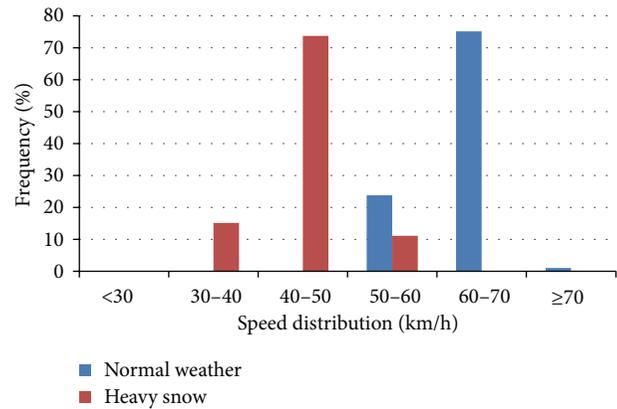


FIGURE 1: Spot speed distribution difference under the same traffic volume levels.

3.2. Travel Speed under Same Traffic Volume Levels. This study selected certain road sections which typically have the same traffic volume levels and similar geometric design to analyze the traffic data collected separately during normal weather and snowy weather conditions. Hence, after the velocity distribution analysis based on the data which was collected on January 3, 2010 with heavy snow and December 27, 2009 with normal weather during the time period of 6:00 AM to 10:00 PM, travel speed reductions caused by the heavy snow were revealed. The influence of various weather conditions on speed distribution is summarized in Table 3.

As can be seen from Table 3 and Figure 1, the spot speed distributions under the same traffic volume levels have obvious differences due to the impact of heavy snow. The average speeds generally drop by 10–20 km/h with heavy snow. During normal weather, the speed mainly is distributed in the range of 60–70 km/h, which accounts for about 70% in all collected samples. Under heavy snow conditions, the speed was mainly distributed in the range of 40–50 km/h. This decreased speed was caused by the road conditions and driver's behaviors when there was heavy snow. The road surface and markings were obscured by the snow, and the visual fields were limited. Accordingly, drivers subconsciously decreased their speeds as they became used to the surrounding environment. Thus, the peak values of speed under the snow conditions were much lower than normal.

4. Speed-Volume Relationship Analysis

The observed traffic volume on expressways changes along with the travel demand and the road network operation level.

TABLE 2: Traffic flow characteristics of expressway under various weather conditions.

Time period Weather condition	Morning peak		Evening peak		Off-peak	
	Heavy snow	Normal weather	Heavy snow	Normal weather	Heavy snow	Normal weather
Average speed (km/h)	56.60	54.96	59.84	46.05	64.36	71.71
Volume per hour (veh./h)	1632	4722	1378	4952	832	2751
Speed change on heavy snowy day (%)	+2.98		+29.95		-10.26	
Volume reduction (%)	65.44		72.17		69.76	

TABLE 3: Spot speed distribution in certain expressway section.

Spot speed (km/h)	Distribution					
	<30	30-40	40-50	50-60	60-70	≥70
Normal weather	0	0	0	23.81	75.15	1.04
Heavy snow	0	15.14	73.72	11.14	0	0

The relationship between the observed flow rate and the velocity can reveal the characteristics of expressways under specific weather conditions and the road environment. The parameters including free-flow speed and saturation flow are usually used to reflect the difference. The research illustrates these data in the speed-flow rate scatter diagram by using the traffic flow data collected under normal weather and heavy snowy weather conditions (see Figure 2) and also conducted the preliminary regression fitting.

It can be clearly seen from the scatter diagram in Figure 2 that the maximum hour flow rate in the normal weather situation is much larger than the one in the heavy snow condition, and the average free-flow speed under heavy snow is about 20 km/h lower compared to those under normal weather. The vertex of the flow rate—the velocity curve corresponding to the flow rate—can be theoretically regarded as the saturation flow which is also the capacity of a single lane [22, 23]. As shown in the scatter diagram, the capacity of a single lane on the expressway is approximately 1800 vph in normal weather, but the snow caused the curve to show inward side displacement, and the single lane capacity is about 1200 vph, which is a remarkable 33% drop. It also means that the saturated vehicle headway in heavy snow is much higher than that of normal weather. The obvious differences are caused by the traffic volume decrease, and the drivers' behavior changes under snow conditions. Because the braking performance of vehicles on the slippery road surface shows a noticeable decline during heavy snow, the drivers tend to take action to modify speeds and keep greater distances between vehicles.

The horizontal line passes the vertex of the relationship curve dividing the curve into two parts with an upper half curve and a lower half curve. Thus, the corresponding speed of this line is defined as the speed at capacity [22]. The traffic flow among the region encircled by the level line and the upper half curve is a steady flow, and the vehicles can drive at a steady and smooth speed. Conversely, the region encircled by the level line and the lower half curve is the force flow area, and the traffic flow in this region has to move at a low speed near the congestion condition.

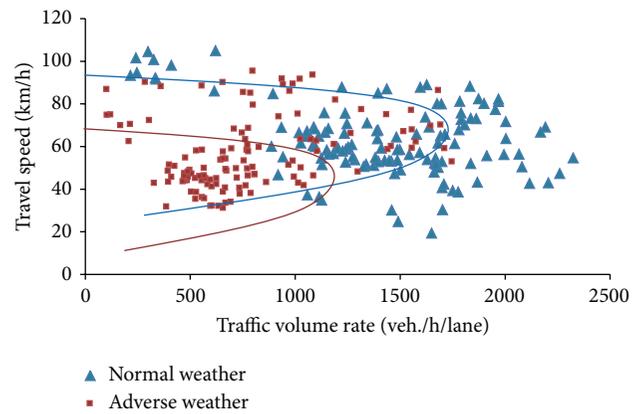


FIGURE 2: Observed speed-flow rate scatter diagram under various weathers. Note: the two regression lines in the figure are just the sketch fitting.

As seen in Figure 2, the majority of the data points are distributed in the upper part under normal weather, and the vehicles can drive at a higher speed than the speed at capacity. However, the majority of the data points are distributed on both sides of the critical speed in the heavy snow situation, and partial vehicles were influenced by the snow and were forced to drive at a lower speed.

5. Reduction Coefficients of Traffic Flow under Various Snow Densities

From the former analysis, it can be concluded that snow has an obvious impact on traffic flow parameters on the expressway and often leads to various degrees of speed reductions. However, there are different reduction percentages under different snow densities. Consequently, besides the heavy snowy day (January 3, 2010) and normal weather day (December 27, 2009), we also collected data from the same detectors on November 10, 2009 with medium snow (snow precipitation is about 3.0 mm) and November 17, 2009 with

TABLE 4: Reduction coefficients of traffic flow parameters under various snow densities.

Snow density		Heavy snowy day	Day after heavy snow	Moderate snowy day	Light snowy day
Average	Volume	42.70	39.05	13.95	2.11
Reduction percentage (%)	Speed	28.01	5.07	6.60	1.77

normal weather as well as on January 9, 2010 with light snow (snow precipitation is about 0.7 mm) and January 16, 2010 with normal weather. The study analyzed the detector data collected from the expressway on several snow days with various snow densities, and the reduction coefficients of traffic flow parameters under light snow, medium snow, and heavy snow were derived from the analysis.

5.1. Reduction Coefficient under Heavy Snow

5.1.1. The Snow Day. Beijing experienced a snowstorm day on January 3, 2010, and the thickness of snow exceeded 20 cm. The data on December 27, 2009 with normal weather was selected to compare with the data on the snowy day. The same volume level data on both days were selected to analyze the spot speed reductions.

Because the observed snowy day was on the weekend, rigid travel demand was small; thus, the entire day's travel demand substantially dropped as a result of the heavy snow, but there were different degrees of reductions in different locations. The flow rates in the partial data set had a decrease of 40%–60% which indicated that these expressway road sections had a high proportion of elastic travel demand, and the road section users were highly sensitive to the weather conditions. On the other hand, the other partial observation points had a volume decline of about 15%~25% which means that the travel demand on these expressway road sections had strong rigidity. The mean traffic volume reduction was about 40%, which is a little smaller than the former road sections.

In addition, the researcher chose multiple groups of data under similar traffic flow levels on both days to compare the speed drops under the heavy snow conditions using contrast analysis. The data analysis indicated the following conclusions: (1) under the different traffic flow levels, the degree of speed reductions was different; (2) the lower the volume level, the greater the speed reduction; (3) the speeds declined at a percentage between 15% and 40%, and the average speed drop under heavy snow was about 28%.

5.1.2. The Subsequent Day after Snow. The data collected on January 4, 2010 was selected to illustrate the traffic performance in the next day after heavy snow. Some roads were icy, and the road conditions had worsened further on that day. As a result of the working day, the majority of travelers had rigid travel demands to go to work, but more people are attempt to travel with public transit; therefore, the traffic volume reduction was slightly smaller than the weekend at 20%–50%. There were almost no obvious differences during the peak hours and the off-peak hours with a volume reduction for both at about 40%. At the same time, the driving speed did

not obviously drop because the volume on the expressway showed a noticeable decline compared to the workday in normal weather. This indicates that the traffic controls and the policies have effectively guaranteed the road network movement and maintained an acceptable service level.

Based on the former data analysis, we can assume that the traffic flow characteristics on the heavy snowy day have a remarkable impact, but traffic conditions of the expressway on the subsequent day are not obviously worse with the help of traffic management countermeasures and the implementation of the adjustable work time policy. The work time is flexible as suggested by municipality, and the pupils are suspended from school for safety, and these countermeasures objectivity reduced the travel demand on that day. The average speed on the heavy snowy day dropped about 28%, and the speed on the next day only had an approximate 5% reduction. This indicates that Beijing has suitable traffic controls, and the organization measures have brought about a positive outcome to alleviate the impact of heavy snow.

5.2. Reduction Coefficient under Lighter Snow. To reveal the different influences on expressway traffic flow, data collected on days with various snow densities between November, 2009 and January, 2010 were also analyzed in this paper. Based on the data analysis, the results are concluded as follows: (1) under heavy snow (precipitation is greater than 5.0 mm), drivers will reduce their travel demand, and the snow coverage and visibility will influence traffic flow and expressway operations and often lead to an obvious reduction in both traffic volume and travel speed; (2) with moderate snowfall (precipitation is greater than 2.5 mm and smaller than 5.0 mm) and light snow (precipitation is smaller than 2.5 mm), the impact of the snow on expressway traffic flow is small, and the volume and the speed only have slight reductions; (3) Table 4 shows the reduction coefficients of traffic flow parameters under various snow densities.

In conclusion, the reduction coefficients of traffic flow parameters under various snow densities in Beijing can be recommended. The volume reductions on the expressways separately are 40%, 15%, and 5% under heavy snow, moderate snow, and light snow, respectively, and the average speed of expressway reductions is separately 25%, 8%, and 0% under the heavy snow, moderate snow, and light snow, respectively.

6. Conclusions

Based on comparisons between traffic flow data and parameters under various snow intensities on snowy days and normal weather days in different snowfall precipitation day, the paper recorded the reductions of traffic flow parameters

on the expressway. The results indicate that the average speeds on the expressway under heavy snow conditions have a 10–20 km/h decrease when compared to those under normal weather conditions, the saturated vehicle headway generally increases, and the road capacity drops about 33%. Furthermore, a specific expressway traffic parameter reduction model under various snow volume levels was developed. It summarized the reduction coefficients of expressway volumes and speeds at same volume levels under various snow densities in Beijing. The paper's conclusions provide an effective support and reference for urban expressway controls and traffic management strategies to reduce speeds and alleviate congestion during snow conditions.

Nevertheless, this paper is mainly based on the detector data, and the multisource traffic data including video data and loop detectors data can be introduced to improve the accuracy of the reduction model in the future research, or under other different adverse weather conditions, and more microscopic characteristics of traffic flow model can be revealed by using the multisource traffic data.

Acknowledgments

This paper was supported in part by the National Natural Science Foundation of China with no. 51108013 and the Ministry of Industry and Information Technology of China under the Major Program of National Science and Technology with no. 2013ZX01045003-002. The authors would like to show great appreciation for these supports.

References

- [1] J. C. Tanner, "Effect of weather on traffic flow," *Nature*, vol. 169, no. 4290, p. 107, 1952.
- [2] J. C. McBride, "Economic impact of highway snow and ice control," Final Report, Federal Highway Administration Report FHWA-RD-77-95, Washington, DC, USA, 1977.
- [3] Kleitsch and Cleveland, "The effect of rainfall on freeway capacity," Tech. Rep. HSRI TrS-6, Highway Safety Research Institute, University of Michigan, Ann Arbor, Mich, USA, 1971.
- [4] E. R. Jones and M. E. Goolsby, "Effect of rain on freeway capacity," Research Record 14-23, Texas Transportation Institute, Texas A&M University, College Station, Tex, USA, 1969.
- [5] *Highway Capacity Manual Fourth Edition (HCM, 2000)*, Transportation Research Board of the National Academies, Washington, DC, USA, 2000.
- [6] H. Y. Wang, Q. J. Xiang, J. Lu, and Y. W. Ren, "Influence of adverse weather on vehicle trip at freeway," *Journal of Traffic and Transportation Engineering*, vol. 5, no. 1, pp. 124–126, 2005.
- [7] K. Y. Dong and B. C. Lu, "Influences of bad weather to movements of vehicles," *Journal of Chongqing Jiaotong University*, vol. 8, no. 6, pp. 24–26, 2008.
- [8] Q. Li and F. S. Liu, "A moderate expressway vehicle flow control under adverse weather condition," *Shandong Transportation Technology*, vol. 3, pp. 9–11, 2009.
- [9] W. H. Wang, F. G. Hou, H. C. Tan, and H. Bubb, "A framework for function allocations in intelligent driver interface design for comfort and safety," *International Journal of Computational Intelligence Systems*, vol. 3, no. 5, pp. 531–541, 2010.
- [10] "Empirical studies on traffic flow in inclement weather," Federal Highway Administration Report FHWA HOP-07-073, Cambridge Systematics, Inc, and Virginia Tech Transportation Institute, 2006.
- [11] P. J. Maki, "Adverse weather traffic signal timing," in *Proceedings of the 69th Annual Meeting of the Institute of Transportation Engineers*, Las Vegas, Nev, USA, 1999.
- [12] H. C. Lieu and S. M. Lin, "Benefit assessment of implementing weather-specific signal timing plans by using CORSIM," *Transportation Research Record*, no. 1867, pp. 202–209, 2004.
- [13] L. Zhang, P. Holm, and J. Colyar, "Identifying and assessing key weather-related parameters and their impacts on traffic operations using simulation," Federal Highway Administration Report FHWA-HRT-04-131, 2004.
- [14] M. W. Hu, "Impact analysis of inclement weather on traffic signal timing," *Transportation Standardization*, vol. 169, pp. 156–160, 2007.
- [15] 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.
- [16] A. T. Ibrahim and F. L. Hall, "Effect of adverse weather conditions on speed-flow-occupancy relationships," *Transportation Research Record*, no. 1457, pp. 184–191, 1994.
- [17] Y. Y. Ren, *Study on the Characteristics of Traffic Stream and Management Countermeasure at Urban Road under the Condition [M.S. thesis]*, Jinlin University, Jilin, China, 2008.
- [18] A. Al-Kaisy and Z. Freedman, "Weather-responsive signal timing: practical guidelines," *Transportation Research Record*, no. 1978, pp. 49–60, 2006.
- [19] P. R. Boyce, *Lighting For Driving: Roads, Vehicles, Signs, and Signals*, CRC Press, Boca Raton, Fla, USA, 2009.
- [20] S. M. Chin, O. Franzese, D. L. Greene, H. L. Hwang, and R. C. Gibson, "Temporary loss of highway capacity and impact on performance: phase 2," Tech. Rep. ORNL/TM-2004/209, Oak Ridge National Laboratory, 2004.
- [21] L. Y. He, *Traffic Congestion Issues of Urban Arterial Road under Inclement Weather and Countermeasure [M.S. thesis]*, Central South University, Changsha, China, 2009.
- [22] F. T. Ren, X. M. Liu, and J. Rong, *Traffic Engineering*, Communications Press, Beijing, China, 2008.
- [23] X. Q. Zhang, S. An, and H. F. Sheng, "Discrete dynamic road network capacity under adverse weather," *Journal of Harbin Institute of Technology*, vol. 41, no. 7, pp. 85–88, 2009.

Research Article

Optimal Energy Control Strategy Design for a Hybrid Electric Vehicle

Yuan Zou, Hou Shi-jie, Li Dong-ge, Gao Wei, and Xiao-song Hu

National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Yuan Zou; zouyuan@bit.edu.cn

Received 8 September 2012; Revised 3 December 2012; Accepted 25 December 2012

Academic Editor: Geert Wets

Copyright © 2013 Yuan Zou 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 heavy-duty parallel hybrid electric truck is modeled, and its optimal energy control is studied in this paper. The fundamental architecture of the parallel hybrid electric truck is modeled feed-forwardly, together with necessary dynamic features of subsystem or components. Dynamic programming (DP) technique is adopted to find the optimal control strategy including the gear-shifting sequence and the power split between the engine and the motor subject to a battery SOC-sustaining constraint. Improved control rules are extracted from the DP-based control solution, forming near-optimal control strategies. Simulation results demonstrate that a significant improvement on the fuel economy can be achieved in the heavy-duty vehicle cycle from the natural driving statistics.

1. Introduction

Commercial transport vehicles, especially heavy-duty trucks, play an important role in the socialism construction of China. More than a half of the freight transported in China is carried by heavy-duty trucks. The increasing number of merely gasoline/diesel-propelled trucks brings some negative impacts, such as excessive fuel consumption and severe air pollution. In order to reduce the dependence on carbon-based fuel and poisonous emissions, hybrid powertrain has been widely studied recently. Owing to the dual-power-source nature, the complex configuration, and the operation modes, the control strategy of hybrid electric vehicle (HEV) is typically more complicated than that of traditional engine-based vehicle. Therefore, system-level vehicle simulation methodology is often applied to implement accurate sizing and matching studies, as well as to develop effective energy control algorithms, before the final design and physical prototyping.

The existing energy management and control strategies of hybrid vehicle can be mainly classified into three categories. The first type employs heuristic control techniques, such as rules/fuzzy logic for the control algorithm development [1, 2]. The principle is commonly based on the concept of “load leveling,” which attempts to operate the internal combustion

engine in an efficient area and the battery as a load-leveling device used to provide the remaining power demand.

The second type of approach is based on static optimization methods that instantaneously determine the efficient power split between different energy sources by minimizing a cost function. To calculate the cost of energy, the electric energy is translated into an equivalent amount of fuel [3, 4]. Due to its relatively simple point-wise optimization nature, it is possible to extend the optimization scheme to solve the simultaneous fuel economy and emission optimization problem [5].

The fundamental mechanism of the third type of HEV control algorithm considers the dynamic nature of the vehicle system when performing an optimization [6–8]. A time horizon is involved in this type of dynamic optimization, instead of a time instant in the static one mentioned above. Power split algorithms obtained from dynamic optimization are thus more accurate under transient conditions but are computationally more intensive. Despite that dynamic optimization cannot be realized onboard due to its preview nature and heavy computation requirement, it has been recognized as a good benchmark to the first two types of algorithms.

The main purpose of this paper is to use dynamic programming (DP) to solve the optimal control problem of a heavy-duty hybrid truck. We thoroughly analyze and discuss

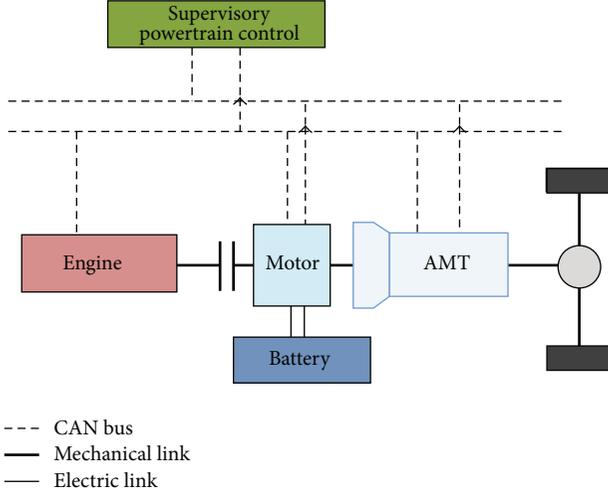


FIGURE 1: Architecture of the hybrid electric truck.

the DP-based results, from which the real-time control rules are extracted. A feed-forward simulation scheme is employed so as to enable the study of vehicle energy control strategy under realistic transient conditions. It can be found that the performance of the developed rule-based algorithm can be improved significantly relative to that based on static optimization.

The remainder of this paper is organized as follows. In Section 2, a heavy-duty hybrid electric truck model is built, followed by an explanation of the preliminary rule-based control strategy. The DP procedure and corresponding optimization problem are introduced in Section 3. In Section 4, the accomplished DP-based results and the real-time control rules are depicted, and the fuel economy results of the developed control strategy are also evaluated. The conclusions are presented in Section 5.

2. Hybrid Electric Truck Model

2.1. Vehicle Configuration. The vehicle architecture is given in Figure 1. The truck has a pretransmission parallel hybrid configuration allowing for a smaller electric motor/generator, an easier packing, and reduced spin losses, compared to the posttransmission type [9]. The diesel engine WP7.210 was manufactured by Weichai Power Co., Ltd. In order to ensure the accessibility of the total peak power, a 90 kW electric motor/generator was selected, and a 60 Ah lithium-ion battery was chosen as the onboard energy storage. A 9-gear-automated mechanical transmission (AMT) was used. The parameters of the vehicle and its components are listed in Table 1.

A system-level HEV model is developed in the MATLAB/Simulink environment, as shown in Figure 2. The HEV model is an accurate but moderate model appropriate for efficient evaluation of fuel economy [10]. Some fast dynamics, like intake manifold filling and motor dynamics, are much quicker than energy dynamics and are thus neglected.

TABLE 1: Parameters of the hybrid electric truck.

DI Diesel Engine	7.0 L, 155 kw at 2000 rpm, 900 Nm at 1300–1600 rpm
Motor/Generator	Maximum power: 90 kw Maximum torque: 600 N-m Maximum speed: 2400 rpm
Lithium-ion Battery	Capacity: 60 Ah Number of modules: 25 Nominal voltage: 12.5 (volts/module)
AMT	9 speed, GR: 12.11/8.08/5.93/4.42/3.36/2.41/1.76/1.32/1
Vehicle	Curb weight: 16000 kg

2.2. Preliminary Rule-Based Control Strategy. The preliminary rule-based control strategy is designed based on the static optimization algorithm minimizing the total equivalent fuel consumption—sum of the fuel consumption and battery energy consumption rates at every step. The equivalent fuel consumption cost is defined as [11]

$$m_{f_total} = m_{f_eng} + m_{f_elec}, \quad (1)$$

where m_{f_eng} and m_{f_elec} represent the fuel consumptions of the engine and electric machines, respectively. m_{f_elec} can be calculated using the following equation:

$$m_{f_elec} = \lambda \cdot \Delta t \cdot \frac{P_{elec}}{\eta_{trans}}, \quad (2)$$

where λ is a conversion factor from electrical energy to engine fuel consumption, and its value was specified as $3.3e - 5$. $\Delta t = 1$ s is the time step. P_{elec} represents the power of electric machines. $\eta_{trans} = \eta_{batt} \cdot \eta_i \cdot \eta_{MG}$ is the total efficiency of the electrical system composed of the battery, inverter, and motor/generator. The optimal control $u(k)$ can be obtained by solving the following equation:

$$\text{Minimize } J = \int_{u(k)} m_{f_total} dt. \quad (3)$$

Here, the control variables include the engine throttle and the transmission gear number, which can be solved using the following function written in the form of .m file in MATLAB

$$[\text{throt_opt}, \text{gear_opt}] = \text{optimal}(\text{shf_spd}, \text{req_power}), \quad (4)$$

where throt_opt and gear_opt are the optimal engine throttle and transmission gear, respectively, shf_spd is the final drive input shaft speed, and req_power is the power requested.

2.3. Fuel Economy Evaluation. A nature driving schedule proposed by Weichai Power is shown in Figure 3. The battery SOC correction procedure proposed in [12, 13] was used to correct fuel economy, when the initial and final battery SOC values are not identical. The hybrid electric truck, given the preliminary rule-based control strategy, achieved a fuel economy of 6.78 miles per gallon (MPG). However, this rule-based control strategy is component based rather than system based. Hence, this control strategy is not globally optimal. This motivates the use of DP as an analysis and design tool.

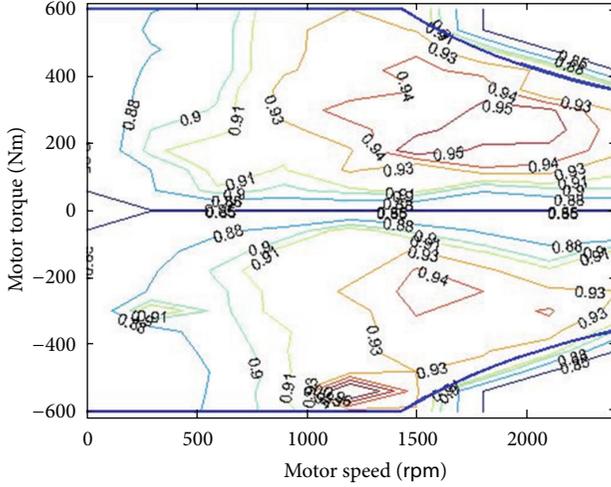


FIGURE 5: Motor efficiency map.

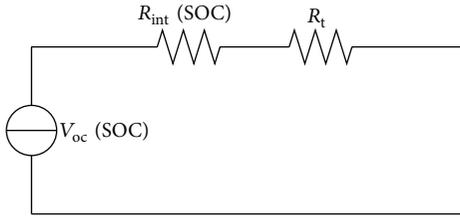


FIGURE 6: Internal resistive battery model.

3.1.2. *Motor.* Similar to the engine, the motor model was established using the experimental data. The motor efficiency is a function of motor torque and speed $\eta_m = f(T_m, \omega_m)$, as shown in Figure 5. Considering the battery power and motor torque limit, the final motor torque becomes

$$T_m = \begin{cases} \min(T_{m,\text{req}}, T_{m,\text{dis}}(\omega_m)), & \text{if } T_{m,\text{req}} > 0, \\ \max(T_{m,\text{req}}, T_{m,\text{chg}}(\omega_m)), & \text{if } T_{m,\text{req}} < 0, \end{cases} \quad (5)$$

where $T_{m,\text{req}}$ is the requested motor torque; $T_{m,\text{dis}}$ and $T_{m,\text{chg}}$ are the maximum motor torques in the motoring and charging modes, respectively.

3.1.3. *Battery.* A 20kWh battery was selected for the target truck, and the equivalent circuit battery model from ADVISOR package was adopted. The schematic diagram of the equivalent circuit model is shown in Figure 6. We ignored the thermal effects and transients so that the only state in the battery model is the SOC as follows:

$$\begin{aligned} \text{SOC}(k+1) &= \text{SOC}(k) \\ &- \left(V_{\text{oc}} - \sqrt{V_{\text{oc}}^2 - 4(R_{\text{int}} + R_t) \cdot T_m \cdot \omega_m \cdot \eta_m^{-\text{sgn}(T_m)}} \right) \\ &\times (2(R_{\text{int}} + R_t) \cdot Q_{\text{max}})^{-1}, \end{aligned} \quad (6)$$

where both the internal resistance R_{int} and the open circuit voltage V_{oc} are a function of the battery SOC. Q_{max} is the maximum battery charge, and R_t is the terminal resistance.

3.1.4. *Driveline.* The driveline is defined as the system from the transmission input shaft to the wheel. The following equations are used to describe the transmission and final drive gear models:

$$T_{\text{wheel}} = \eta_{\text{gear}} \eta_{\text{FD}} \times i_g \times i_0 \times T_i - c_{\text{tm}} \omega_i, \quad (7)$$

$$\omega_i = i_g \cdot i_0 \cdot \omega_{\text{wheel}},$$

where i_g is the transmission gear ratio, i_0 is the final drive gear ratio, and η_{gear} and η_{FD} are the transmission and final drive efficiencies, respectively. T_i is the transmission input torque, c_{tm} is the transmission viscous-loss coefficient, ω_i is the transmission input speed, and ω_{wheel} is the wheel speed.

The gear-shifting sequence of the AMT is simulated using a discrete-time dynamic model:

$$\begin{aligned} \text{gear}(k+1) &= \begin{cases} 9, & \text{gear}(k) + \text{shift}(k) > 9, \\ 1, & \text{gear}(k) + \text{shift}(k) < 1, \\ \text{gear}(k) + \text{shift}(k), & \text{otherwise,} \end{cases} \end{aligned} \quad (8)$$

where the state gear is the gear number, and the control shift to the transmission is restricted to take on the values -1 , 0 , and 1 , corresponding to downshift, sustaining, and up-shift, respectively.

3.1.5. *Vehicle Dynamics.* It is a common practice that only the vehicle longitudinal dynamics are considered. The longitudinal vehicle dynamics are modeled as a point mass to which various forces are applied:

$$\omega_{\text{wheel}}(k+1) = \omega_{\text{wheel}}(k) + \frac{T_{\text{wheel}} - T_{\text{brake}} - r_w (F_r + F_a)}{M_r r_w^2}, \quad (9)$$

where T_{brake} is the friction brake torque. F_r and F_a are the rolling resistance and aerodynamic drag forces, respectively. r_w is the dynamic tire radius, $M_r = M_v + J_r / r_w^2$ is the effective mass of the vehicle, and J_r is the equivalent moment of inertia of the rotating components in the vehicle.

3.2. *DP Problem Formulation.* For an optimization problem, choose $u(k)$ to minimize the cost function

$$J = G_N(x(N)) + \sum_{k=0}^{N-1} L_k(x(k), u(k), w(k)), \quad (10)$$

where

$$x(k+1) = f(x(k), u(k), w(k)), \quad k = 0, 1, \dots, N-1, \quad (11)$$

subject to

$$x(k) \in X(k) \subset \mathfrak{R}^n, \quad u(k) \in U(x(k), k) \subset \mathfrak{R}^m. \quad (12)$$

Here, $x(k)$ is the state vector at the stage k in the space of $X(k)$, which contains the transmission gear number $\text{gear}(k)$ and battery SOC $\text{SOC}(k)$; $u(k)$ is the control vector, including the engine throttle $\text{throt}(k)$ and the transmission gear shift command $\text{shift}(k)$; $w(k)$ is a predetermined disturbance, here it is the rotational speed of the wheel determined by the driving schedule; f is the transition function that represents the system dynamics; L is the instantaneous transition cost; G_N is the cost at the final stage N .

In this paper, the sampling time for this control problem is 1 second. The cost function to be minimized has the following form:

$$J = \sum_{k=0}^{N-1} [L_{\text{fuel}}(k) + \beta |\text{shift}(k)|] + G_N(x_{\text{SOC}}(N)), \quad (13)$$

where $L_{\text{fuel}}(k)$ is the instantaneous cost of fuel use, $\beta |\text{shift}(k)|$ is adjusted to constrain the vehicle drivability avoiding excessive shifting, and β is a positive weighting factor. If there is no constraint on the terminal SOC, the optimization algorithm tends to deplete the battery in order to attain the minimal fuel consumption. Hence, a terminal constraint on SOC $G_N(x_{\text{SOC}}(N))$ is incorporated into the cost function. During the optimization, it is necessary to impose the following inequality constraints to ensure safe/smooth operation of the engine/battery/motor:

$$\begin{aligned} \omega_{e.\min} &\leq \omega_e(k) \leq \omega_{e.\max} \\ \text{SOC}_{\min} &\leq \text{SOC}(k) \leq \text{SOC}_{\max} \\ T_{e.\min}(\omega_e(k)) &\leq T_e(k) \leq T_{e.\max}(\omega_e(k)) \\ T_{m.\min}(\omega_m(k)) &\leq T_m(k) \leq T_{m.\max}(\omega_m(k)), \end{aligned} \quad (14)$$

where ω_e is the engine speed, SOC is the battery state of charge, and SOC_{\min} and SOC_{\max} were selected to be 0.4 and 0.8. T_e and T_m are the engine and motor torques, respectively.

The DP technique is based on Bellman's Principle of Optimality, which states that the optimal policy can be obtained if we first solve a one-stage subproblem involving only the last stage and then gradually extend to subproblems involving the last two stages, last three stages, and so forth until the entire problem is solved. In this manner, the global dynamic optimization problem can be decomposed into a sequence of the simple minimization problems as follows [15].

Step $N - 1$,

$$\begin{aligned} J_{N-1}^*(x(N-1)) \\ = \min_{u(N-1)} [L(x(N-1), u(N-1)) + G(x(N))]. \end{aligned} \quad (15)$$

Step k , for $0 \leq k < N - 1$

$$J_k^*(x(k)) = \min_{u(k)} [L(x(k), u(k)) + J_{k+1}^*(x(k+1))], \quad (16)$$

where $J_k^*(x(k))$ is the optimal cost-to-go function at the state $x(k)$ starting from the time stage k . It represents the optimal resulting cost if the system evolution follows the optimal control law with the current state $x(k)$ from the stage k to the final stage.

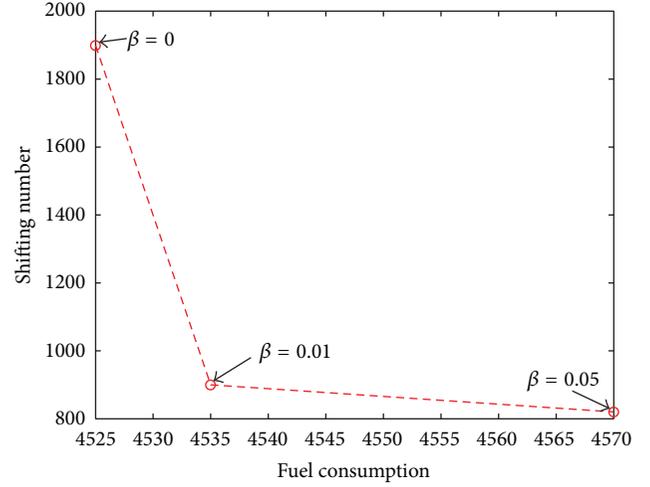


FIGURE 7: Gear-shifting number versus fuel consumption.

The foregoing recursive equations are firstly solved backward and then searched forward to find the optimal control policy. The minimization is performed subject to the inequality constraints shown in (14).

4. Development of Improved Control Strategy

4.1. DP Results. The DP procedure described above produces an optimal solution with respect to (13), where the shift weighting factor β was selected to constrain the AMT shift number. It can be found that a larger value of β results in less frequent gear shifting, but the larger fuel consumption. The balance relationship of the gear shifting number and the fuel consumption is shown in Figure 7, and $\beta = 0.01$ is chosen for the subsequent analysis.

Simulation results of the vehicle based on the DP optimization are shown in Figure 8. The AMT shifts between the 4th and the 9th gear ratio mostly. The SOC increases near 0.6 at the end of the driving cycle due to the penalty in the cost function shown in (13). The power of the engine increases after 1900 s apparently to make SOC a rise. It can be seen that most of the engine operating points fall in the high-efficiency area, near the optimal fuel consumption line before 1400 rpm and near the maximum torque line after 1400 rpm. Compared to the preliminary rule-based control strategy, the fuel economy under the DP optimal control law is improved by 15.9%.

4.2. Development of Improved Control Strategy. The DP control policy is not applicable in real driving conditions in that it requires knowledge of the future speed and load profiles. However, analyzing carefully the DP-based results is able to help us in improving the preliminary rule-based control strategy.

Firstly, we study how the power split based on the preliminary control strategy can be improved. A power-split ratio $\text{PSR} = P_{\text{eng}}/P_{\text{req}}$ is defined to quantify the positive power flow in the powertrain, where P_{eng} is the engine power and

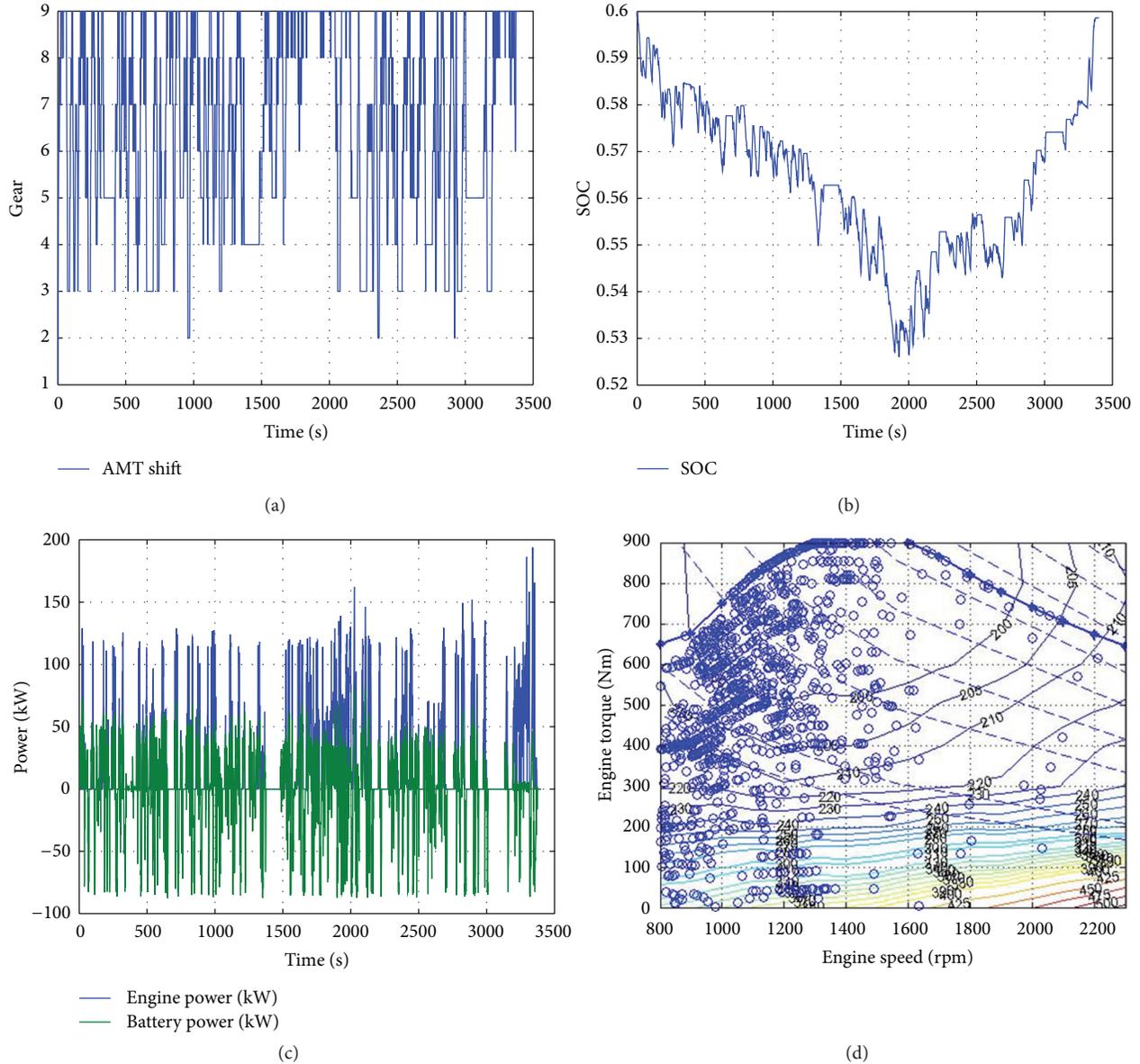


FIGURE 8: DP-based results.

P_{req} is the power requested. Four positive-power operating modes are defined:

- (1) PSR = 0: motor-only mode;
- (2) PSR = 1: engine-only mode;
- (3) $0 < PSR < 1$: power-assist mode;
- (4) PSR > 1: recharging mode.

The enhanced rule can be found by plotting the optimal PSR as a function of the torque demand at transmission input, as shown in Figure 9. It is remarkable that the optimal policy uses the recharging mode in the low torque region, the engine-only mode in the middle torque region, and the power-assist mode in the high torque region. In order to

extract a realizable rule, the least-square curve fit is deployed to approximate the optimal PSR points. The red line in Figure 8 shows the fitting result.

The gear shifting schedule is crucial to the fuel economy of hybrid electric vehicles. In the DP scheme, the gear-shift command is one of the control variables. The gear operating points from the DP-based result are plotted in the form of the power demand versus vehicle speed (see Figure 10). It can be seen that the gear positions are separated into different regions, and the boundaries between adjacent regions represents the optimal gear shifting thresholds. The gear shifting schedule can thus be obtained on the basis of these thresholds.

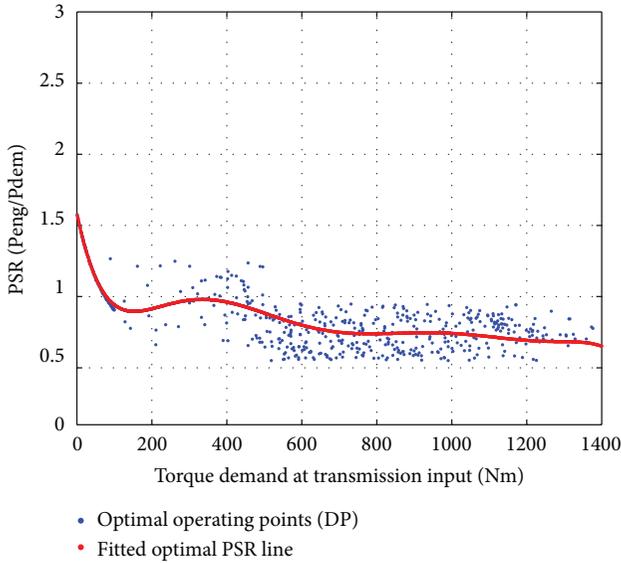


FIGURE 9: DP-based power split behavior.

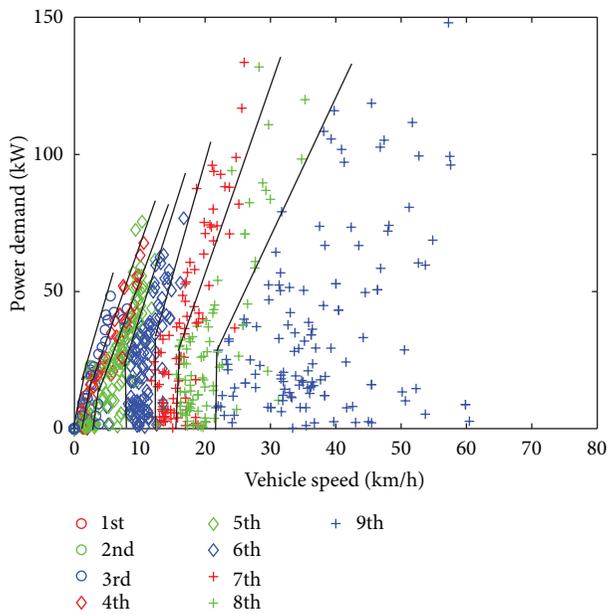


FIGURE 10: DP-based gear operating points.

According to the extracted rules, the improved rule-based control strategy is designed and implemented in the MATLAB/Stateflow environment.

4.3. Fuel Economy Evaluation. The preliminary and improved rule-based control strategies, as well as the benchmark, the DP control policy, are compared by means of the nature driving schedule from the Weichai Power Company Limited. The results are shown in Table 2. It is clear that the improved rule-based control strategy extracted from the DP-based results can ensure a much better fuel economy for the hybrid electric truck, compared with the preliminary one.

TABLE 2: Fuel economy comparison.

	Fuel economy (mile per US gallon)	Improvement
Preliminary rule based	6.78	—
Improved rule based	7.45	9.9%
Hybrid truck (DP)	7.86	15.9%

5. Conclusions

A feed-forward model of a heavy-duty parallel hybrid electric truck is established for studying the optimal energy management strategy. The preliminary rule-based control strategy is designed based on the static optimization algorithm. Based on the simplified energy control strategy, DP is applied to solve the globally optimal energy control strategy. Improved rule-based control rules that are applicable in real time are extracted from analyzing the DP-based results. Based on a natural driving schedule from Weichai Power Company Limited, a comparison is conducted between the improved rule-based control strategy and the preliminary one derived from static optimization. The result demonstrates that the improved rule-based control policy from the DP algorithm results in a noteworthy enhancement on the fuel economy of the hybrid electric truck. The proper approximation of DP behavior, however, is the key process to extract the control rule and is worth further investigation.

Acknowledgments

The authors would like to express their deep gratitude to Professor Huei Peng in The University of Michigan for many helpful suggestions on the DP algorithm. This work was supported by the Natural Science Foundation of China (50905015).

References

- [1] B. M. Baumann, G. Washington, B. C. Glenn, and G. Rizzoni, "Mechatronic design and control of hybrid electric vehicles," *IEEE/ASME Transactions on Mechatronics*, vol. 5, no. 1, pp. 58–72, 2000.
- [2] N. J. Schouten, M. A. Salman, and N. A. Kheir, "Fuzzy logic control for parallel hybrid vehicles," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 3, pp. 460–468, 2002.
- [3] A. Sciarretta and L. Guzzella, "Control of hybrid electric vehicles," *IEEE Control Systems Magazine*, vol. 27, no. 2, pp. 60–70, 2007.
- [4] G. Paganelli, G. Ercole, A. Brahma, Y. Guezennec, and G. Rizzoni, "A general formulation for the instantaneous control of the power split in charge-sustaining hybrid electric vehicles," in *Proceedings of the 5th International Symposium on Advanced Vehicle Control (AVEC '00)*, Ann Arbor, Minn, USA, 2000.
- [5] V. H. Johnson, K. B. Wipke, and D. J. Rausen, "HEV control strategy for real-time optimization of fuel economy and emissions," in *Proceedings of the Future Car Congress*, SAE paper no. 2000-01-1543.

- [6] A. Brahma, Y. Guezennec, and G. Rizzoni, "Dynamic optimization of mechanical electrical power flow in parallel hybrid electric vehicles," in *Proceedings of the 5th International Symposium on Advanced Vehicle Control (AVEC '00)*, Ann Arbor, Minn, USA, 2000.
- [7] U. Zoelch and D. Schroeder, "Dynamic optimization method for design and rating of the components of a hybrid vehicle," *International Journal of Vehicle Design*, vol. 19, no. 1, pp. 1–13, 1998.
- [8] C. C. Lin, J. M. Kang, J. W. Grizzle, and H. Peng, "Energy management strategy for a parallel hybrid electric truck," in *Proceedings of the American Control Conference*, pp. 2878–2883, Arlington, Va, USA, June 2001.
- [9] J. M. Miller, *Propulsion Systems for Hybrid Vehicles*, The Institution of Engineering and Technology, Stevenage, UK, 2004.
- [10] S.-J. Hou, Y. Zou, and R. Chen, "Feed-forward model development of a hybrid electric truck for power management studies," in *Proceedings of the 2nd Intelligent Control and Information Processing (ICICIP '11)*, Harbin, China, 2011.
- [11] J. Liu, *Modeling, configuration and control optimization of power-split hybrid vehicles [Ph.D. thesis]*, University of Michigan, 2007.
- [12] D. L. McKain, N. N. Clark, T. H. Balon, P. J. Moynihan, S. A. Lynch, and T. C. Webb, "Characterization of emissions from hybrid-electric and conventional transit buses," SAE Paper 2000-01-2011, 2000.
- [13] Society of Automotive Engineers, "Recommended Practice for Measuring Exhaust Emissions and Fuel Economy of Hybrid-Electric Vehicles," Hybrid-Electric Vehicle Test Procedure Task Force SAE J1711, 1998.
- [14] I. Kolmanovsky, M. van Nieuwstadt, and J. Sun, "Optimization of complex powertrain systems for fuel economy and emissions," in *Proceedings of the IEEE International Conference on Control Applications (CCA '99)*, pp. 833–839, August 1999.
- [15] R. Bellman, *Dynamic Programming*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, USA, 2010.

Research Article

A Pilot Study Verifying How the Curve Information Impacts on the Driver Performance with Cognition Model

Xiaohua Zhao, Wei Guan, and Xiaoming Liu

Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Xiaohua Zhao; zhaoxiaohua@bjut.edu.cn

Received 18 September 2012; Revised 27 December 2012; Accepted 28 December 2012

Academic Editor: Wuhong Wang

Copyright © 2013 Xiaohua Zhao 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.

Drivers' misjudgment is a significant issue for the curve safety. It is considered as a more influential factor than other traffic environmental conditions for inducing risk. The research suggested that the cognition theory could explain the process of drivers' behavior at curves. In this simulator experiment, a principle cognition model was built to examine the rationality of this explanation. The core of this pilot study was using one of the driving decision strategies for braking at curves to verify the accuracy of the cognition model fundamentally. Therefore, the experiment designed three treatments of information providing modes. The result of the experiment presented that the warning information about curves in advance can move the position of first braking away from curves. This phenomenon is consistent with the model's inference. Thus, the conclusion of this study indicates that the process of the drivers' behavior at curves can be explained by the cognition theory and represented by cognition model. In addition, the model's characteristics and working parameters can be acquired by doing other research. Then based on the model it can afford the advice for giving the appropriate warning information that may avoid the driver's mistake.

1. Introduction

The safety at curves is an important issue for drivers as well as traffic engineers. Many investigations and researches had reported that a huge number of crashes or accidents happen at curves every year [1, 2]. The causation of this phenomenon is that while the vehicle passing through a curve, the additional centripetal force will exert on the car. That force will induce a difficult driving task required the driver to complete [3], then drivers may easily make a mistake and cause an accident. To handle this problem, many traffic facilities, in-vehicle systems, and standards had been invented (or specified) to improve driving environment and curve safety [4, 5]. Although these measures made curve driving safer, there are still a lot of crashes or accidents caused by drivers. Thus, many researchers have shifted their focus on studying the process of drivers' behavior at curves [6]. Moreover, based on the results/findings of those studies, engineers may improve or modify the measures to promote the curve safety. The

characteristics of curve safety and driving models are two main basic elements of the researches.

Curve safety research takes a large portion of traffic studies. According to researches, a higher vehicle speed may largely increase the probability of crashes, but the direct reason which lead to accidents is always the drivers' incorrect operation [7, 8]. While a vehicle is passing through curves, it requires more attention resources to collect information, more mental sources to make decisions, and more operations to perform as quickly and preciously as possible for drivers. High vehicle speed decreased the required time of the whole working process for drivers, which means drivers have to finish the process more quickly. As a result, the error tolerance decreased, and the accident possibility was increased. Comte and Janso's study also showed that a large proportion of curve accidents were caused by a driver traveling too fast through a curve and then losing control of the vehicle or being forced into a skid and cannot perform the right operation

[9]. In addition, it was a consensus that increasing degrees of curvature caused more accidents [10]. Another study found that the driver's correct expectation to an encountered curve could make driving safer [11].

In all, the key point to improve curve safety was to let the drivers know the information of coming curves well, then they can be attentive to perceive traffic information accurately, make appropriate decisions easily, and be prepared for more driving operations. Specifying the driver's working process clearly was the precondition for providing the appropriate information to drivers.

According to the cognition theory, driving performance is consisted by 3 phases: information perception, driving decision, and operation execution [12]. In addition, both phases of driving decision and operation execution partly rely on the information coming from the phase of information perception. Therefore, the information perception is the basis for the cognition. In this phase, the driver transforms the picture of real world into information that will be used in the other two phases. The more exact the transformed information was, the easier the driver can make appropriate decisions and correct operations.

Although the driver has five key perception systems, the visual system takes a considerable percentage in driving task [13]. Based on the visual system the driver would rebuild a driving scenario in their brain. The scenarios may reflect the past, current, and future driving situations. According to all the three scenarios drivers can decide their driving strategy. Thus, acquiring exact and valid information is the most important guarantee for curve driving safety. However, factors like environment, driver ability, or distraction may cause the acquisition of inexact and invalid information. As studies shown, most collisions were due to drivers' misperception of frontal curves [14] or failing to obey traffic signs [15]. Also, there is a consentaneous finding: while a driver is in the attention status or has received a warning [16], the probability of an accident is prominently reduced. In addition, Roca Javier et al.'s research states that if one wants to finish a driving task well, he must properly perceive traffic objects (e.g., road signs), maintain an appropriate state of alertness to make decisions, and perform the operation at the right time [17].

In summary, the objective of this paper is using the cognition theory to build a principium model to describe curve driving. This model is driven by information, and different information should induce different operation. To verify the accuracy of the cognition model and related assumption, we compared the driving operations with three different information origins. The result indicates that the curve driving behavior can be well explained by cognition model.

2. Cognition Model

Building a cognition model for the curve driving was the primary task in this study. However, the goal of this model was neither to simulate the driver's behavior accurately nor to rebuild the whole working process of the human brain. This

study mainly focuses on representing an equivalent model for curve driving.

2.1. Cognition Theory. Cognition theory derived into two basic subjects: psychology and neurology. The former one focuses on studying the subjective human (just like how human think, why they have self-awareness), and provide a lot of basic theoretical and rich practical materials for cognition; the latter one is the study of objective human existence; through experiments it verifies a large number of assumptions for cognitive principles. Therefore, the cognition study may involve in both the subjective and objective aspects of human, which can give a comprehensive principle frame for explaining human behaviors.

At present, cognition theory has already been applied in several fields of study, and in the case of different studying priorities, the researchers had given the definition respectively from different academic areas. The followings are three representative definitions:

- (1) cognitive architectures are designed to simulate human intelligence in a human-like way [18];
- (2) a scientific hypothesis about those aspects of human cognition those are relatively constant over time and relatively independent of task [19];
- (3) cognition is a term referring to the mental processes involved in gaining knowledge and comprehension, including thinking, judging, remembering, and problem solving [20].

Based on these definitions, this paper argues that the core of cognition theory is "studying how human respond to the objective/subjective things and the correlative characters, rules and principles". In addition, the final goal of cognitive studying is to build an equivalent model to represent how human respond to the objective/subjective things. Furthermore, the two contents of cognition model are cognition architecture and cognition process, as the following specification.

Cognition architecture can be defined as a set of functional modules, which are integrated in a specific order. The research about cognition architecture includes the modules' characters and their relationship. Generally, the modules include attention, memory, problem solving, decision-making, learning, and so on.

Cognition process refers to a temporal logical sequence to accomplish a cognitive task, which is based on specific cognition architecture. Some people claim that knowledge is the foundation of the cognition process [21], while others believe that the cognitive process is driven by events [22]. This paper assumes that information is the fundamental element in cognition process. Because every subjective and objective thing had uniform and inherent property, that is, information. In other words, everything is either a subjective or an objective representation of information. For example, the real curve warning sign on the roadway can be seen as objective representation of information of "there is a curve ahead"; yet when a driver knows the meaning of the traffic

sign it converts to subjective reproduction in the driver's mind.

For this assumption, we can unify all kinds of things in the cognition model under a uniform measurement and consider various factors effectively. Before analyzing the factors affecting driving behavior, it can convert traffic sign, road condition, driving knowledge, or other things all into the information, which may help studying.

For this assumption, we can unify all kinds of things in the cognition model under a uniform measurement and consider various factors effectively. Before analyzing the factors affecting driving behavior, it can convert traffic sign, road condition, driving knowledge, or other things all into the information.

2.2. Model for Curve Driving. According to cognition theory, the cognition architecture of this model is composed by five kinds of general modules, involving perception module, transmission module, memory module, process module, and motion module. Each kind of modules has its own characters and function in different cognition models. The modules in this study can be described as follows.

- (i) Perception module: obtain traffic information from the external environment. There are five kinds of perceptions for human: visual, aural, tactual, olfactory, and gustatory. This study only makes use of visual perception for testing cognition model.
- (ii) Process module: convert the input information to output information. The conversion can be described as productions or functions. The core process in this study is driving decision module. It can convert traffic information to the operation command information basing on the driving strategy. The driving strategy will be specified in Section 2.3.
- (iii) Memory module: this module can be classified into two categories, short-term memory and long-term memory. The short-term memory was employed to store information which the process module's need, and the long-term memory was employed to store different kind knowledge library. This study involves both two kinds of memory modules. Short-term memory includes image buffer, knowledge buffer and constrain buffer, and long-term memory is the knowledge library, seen Figure 1.
- (iv) Motion module: perform the command information coming from driving decision module.
- (v) Transmission module: transmit information between different modules.

Based on modules mentioned above the cognition architecture can be represented as Figure 1.

Figure 1 has four kinds of information channels, and the channels are tagged by different arrows and box styles as the legend shown. The white boxes are the start or the end of the channels and each channel has its own function as described below.

- (i) The perception channel is used to collect information from traffic environment. The information of traffic environment is regarded as an image. Through the visual module and transmit module, it is stored in the image buffer as other analog signals. The visual encoder module converts those analog signals to abstract traffic information through the transmit module, and it will be used in the decision module. (The working process of other perception channels functions in the same way, this study will not focus on them.)
- (ii) The motion channel is used to generate driving operation. When a driver makes a decision, it sends the command information to motor modules to perform through this channel.
- (iii) The learning channel through the learning system transmits information from message buffers into the knowledge library.
- (iv) The memory channel is used to convert the information from knowledge library to short-term memories. In this study, the goal analysis module selects the related knowledge according to the main driving goal, then puts them into constraint buffer and knowledge buffer for the process modules.

Although this architecture is simpler than the real driver's working process, it is still too complex for this pilot experiment to test all parts. It needs a further predigest by giving some constraint. The simplified model is shown in Figure 2.

In this model, the driving decision module has two information sources: one is visual perception; the other is knowledge library. The driving decision module makes a decision in accordance with information from these two sources. Then the motor system should output operation according to the decision. The different information input from two sources can eventually lead to different driving operations. Compared with Figure 1, the module in Figure 2 reduces the other perception channels, encoder module, constraint buffer, and transmit modules. Provided that all modules are maintained at a constant level, the output change is only caused by the two information sources. Therefore, we can simplify Figure 1 to Figure 2.

Based on the architecture, the cognition process is shown in Figure 3. The process starts from the image of traffic environment (seen as input) and ends with driving behavior (seen as output). Through the three phases: information acquiring, decision making, and driving operation, the three processes are executed parallel while the information is serial. It means at a particular moment, the decision-making process is caused by the previous visual perception.

2.3. Model Analysis. Based on the cognition theory, this study argued three assumptions about the cognition model and designed three information providing treatments. Under these conditions, the study makes the inference about the model to verify the rationality with the simulator experiment.

Assumption One: the strategy of driving decision module is that once the information "there is a curve ahead" is

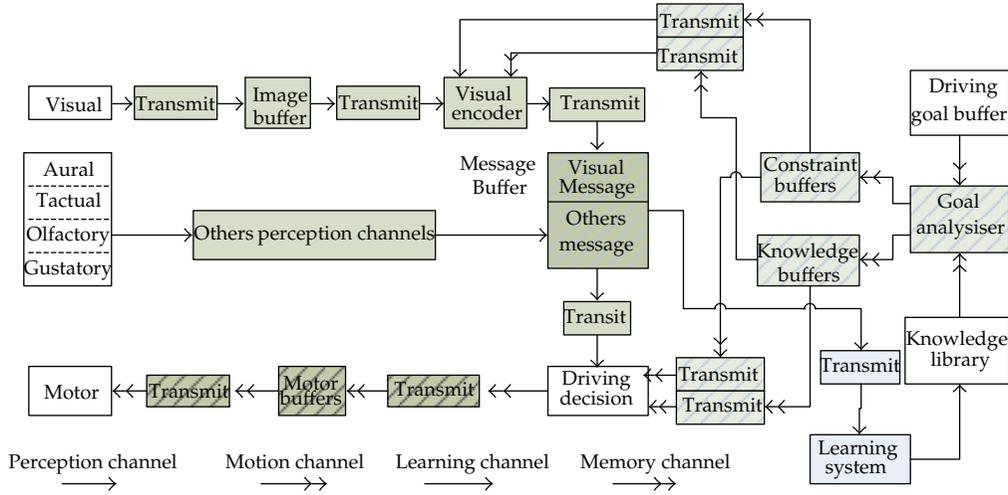


FIGURE 1: The architecture of cognition model.

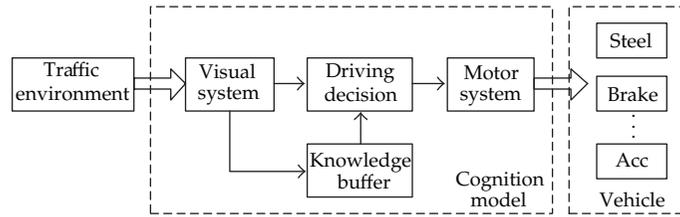


FIGURE 2: The simplified cognition model.

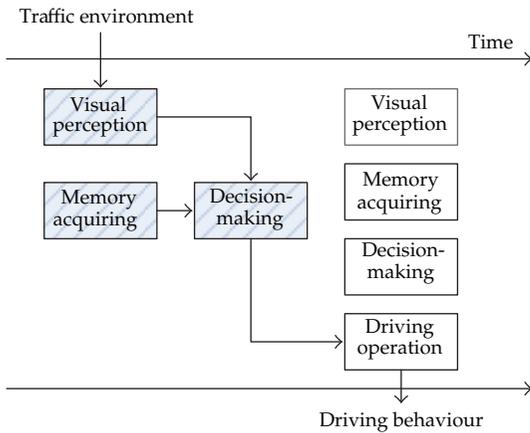


FIGURE 3: The cognition process of the model.

confirmed and vehicle speed is higher than the limitation of passing curve, then the braking operation will be taken.

Assumption Two: the acquired information has a confidence value. If the sum of the confidence values with the same information is over a limen, this information can be confirmed as true.

Assumption Three: information acquired from visual perception channel can be partly stored in the knowledge library via learning system. The stored content may be enhanced by repeating.

Treatment A: the curve information without advanced warning information provided via visual perception channel.

Treatment B: the curve information with advanced warning information provided via visual perception channel.

Treatment C: the curve information provided via visual perception channel, and the advanced warning information provided via memory channel from the knowledge library.

Inference: in the treatment A, the decision module takes more time than other two treatments to confirm the curve information, then the model will brake at last. Since the decision module lacks the warning information, it has the lowest confidence value about the curve information. Furthermore, the higher confidence value may appear to confirm the information and do earlier braking operation.

3. Methods

3.1. Participants. All drivers are at the same level for this study. 18 male participants, having had their license for at least 2 years (average = 4.13, SD = 1.2), aged 22–32 (average = 27.8, SD = 2.24), were recruited from the public. None of the subjects had color vision deficiencies. All of them possess normal or corrected-to-normal vision.

3.2. Apparatus. It is an appropriate choice to use the simulation experiment in this study mainly for three reasons: (1) keeping all trials under the same conditions; (2) making the experiment repeatable; (3) protecting participants from risk.



FIGURE 4: The AutoSim simulator system.

The whole experiment made use of the AutoSim simulator system (see Figure 4) at the laboratory of the Transportation Research Center at Beijing University of Technology.

The hardware is consisted by a vehicle, eight-networked computers (one is the master computer, one communicates with the vehicle system, and the other six are used to compute six different real-time views), a motion control device, and other equipment (such as video and audio devices). Beside, three main softwares used in the simulator experiment: Evariste (used to create experimental scenarios), Simword (for controlling the scenarios), and Scancer (for collecting data and generating car motion). The simulator can record data during the experiment including the action of the accelerator/brake pedal, the steering wheel, vehicle speed, and so forth. The frequency of recording is 30 Hz.

3.3. Experimental Design and Procedures. In order to test the rational of cognition model, the simulator experiment must keep the same constraints with the model declared. The simulator environment can avoid interference or other uncertainties in the actual environment.

First, considering that there is more than one strategy for drivers making choices in driving, the most important constraint to be controlled is to make sure that drivers take the driving strategy, as Assumption one mentioned, when they encounter with the curve. If some other factors did not fit the condition of this strategy, the drivers may not follow that strategy. To access to this goal, the scenarios of this simulator experiment are designed as a single lane of rural roadway with no traffic flow. Considering that the vehicle needs to reach a higher speed than the limitation before entering into the curve, we designed an 800 m straight lane ahead of each curve for drivers' accelerating.

Second, the drivers' ability and driving state should also be under control. According to Figure 1, there are many kinds of factors affecting the driving operation. So when recruiting the participants, all the following characters are considered (age, gender, education, and driving frequency) to avoid these kinds of discomfort. Additionally the operator will give participants an instruction to ensure the driving task for experiment before driving.

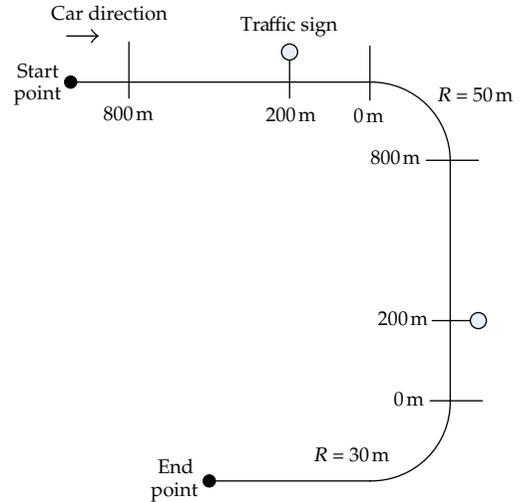


FIGURE 5: Five position of the traffic sign before the curves.

Third, corresponding to three the information providing treatments for the cognition model, 18 participants were divided into three groups randomly in this experiment. In group A, participants drove through curves without traffic sign providing warning information. In group B, participants drove through curves with traffic sign providing warning information. In group C participants drove through curves without traffic sign, but were provided with warning information via the map of scenario before driving. We designed two scenarios (shown in Figure 5), and the difference between them is whether there is a traffic sign before curves. Each scenario has two curves of which the radii are 30 m and 50 m, with a length of 47.1 m and 78.5 m, respectively. The position of traffic sign is 200 m away from the curve.

Finally the procedure of this experiment is stated as follow.

- (i) Filling in basic information to acquire the characters information and driving state about drivers.
- (ii) Participants have had a driving test to avoid drivers' making mistake because they are unfamiliar with the simulator.
- (iii) When the drivers finished the driving test, the operator read the guidance to the driver and gave him the map (in group C).
- (iv) After that, the drivers started the formal experiment, and each participant should drive repeatedly for five times, and there would be a rest for 2-3 minutes between the laps.

4. Result

According to the cognition model, the position of braking start is related with the time when the information "there is a curve ahead" is confirmed. The data of two kinds of braking operations (the first position to release the accelerator pedal, FPRA; the first position to press the brake pedal, FPPB) are used to verify the inference mentioned above.

TABLE 1: The result of braking data of two curves.

Lap 1	FPRA-R30		FPPB-R30		FPRA-R50		FPPB-R50	
	Average	SD	Average	SD	Average	SD	Average	SD
Group-A	104.78	32.19	86.51	34.64	103.89	51.31	72.40	34.26
Group-C	192.10	108.60	117.18	66.95	184.97	57.61	145.36	58.49
Group-B	197.44	28.25	158.94	42.51	185.28	60.17	169.11	87.31
Lap 2-5	Average	SD	Average	SD	Average	SD	Average	SD
Group-A	160.07	58.76	123.49	37.17	188.41	67.10	151.09	69.78
Group-C	201.22	66.27	164.16	46.36	199.28	47.23	166.32	45.24
Group-B	218.80	49.86	188.48	39.44	253.48	72.88	216.59	69.73

TABLE 2: The confidence value of different conditions.

	Curve	Traffic sign	Memory	Total
Group A, lap 1	A1	0	0	A1
Group B, lap 1	A1	B1	0	A1 + B1
Group C, lap 1	A1	0	C1	A1 + C1
Group A, lap 2-5	A1	0	C2	A1 + C2
Group B, lap 2-5	A1	B1	C2	A1 + B1 + C2
Group C, lap 2-5	A1	0	C1 + C2	A1 + C1 + C2

Notice: A1 is a variable. Others are constant. All of them are positive.

First, according to the data in Table 1, following results can be reached.

- (1) The data of group A in lap 1 have the lowest value.
- (2) In laps 2-5, all the four average data (FPRA-R30, FPPB-R30, FPRA-R50, and FPPB-R50) of group A have the lowest value.
- (3) The data of group B in laps 2-5 have larger value than data in lap 1.
- (4) The data of group C in laps 2-5 have larger value than data in lap 1.
- (5) The data of group B have larger value than data of group C in lap 1.
- (6) The data of group B have larger value than data of group C in laps 2-5.

The results may be explained by the cognition theory and fitness of the model. Seen in Table 2, the confidence values of the information “there is a curve ahead” come from different information channels. Besides, the total value of each group has been defined. A1 is the value of drivers acquiring information from visual channel, which is variable, increased by the vehicle travelling near to the curve. B1 is the value of drivers acquiring warning information from traffic sign. C1 is the value of drivers acquiring warning information from drivers’ memory acquired from the map of scenario.

C2 is the value of drivers acquiring warning information from drivers’ memory acquired by learning system through repeated driving.

According to the basic inequalities, we can safely draw a conclusion as below:

- (1) $A1 < A1 + B1, A1 + C1, A1 + C2, A1 + B1 + C2, A1 + C1 + C2$
- (2) $A1 + C2 < A1 + B1 + C2, A1 + C1 + C2$
- (3) $A1 + C1 < A1 + C1 + C2$
- (4) $A1 + B1 < A1 + B1 + C2$

Those inequalities and the cognition assumption can perfectly explain the results of Table 1. As the assumption argued, the braking operation is the result of confirmation of the information “there is a curve ahead”. And to confirm this information the total confidence value must be over limens. Because A1 is a variable increasing with driving time. If a positive value is added, the total value can reach the limens earlier, namely, the position of the drivers beginning braking will be away from the curve. Apparently, the inequalities above explained the result (1) to (4), respectively. Indirectly, the result (5) may infer the inequality “B1 > C1”, which can be verified by result (6).

5. Conclusion

The information is regarded as the bridge between the cognition model and drivers, because both of sides are working based on the information. Studying the information process and transmit channel can obtain how drivers generate driving operation and build a cognition model to represent driving performance. So the information about curve can be seen as the most crucial factor for driving operation. Based on the cognition model and the results of the simulator experiment, this study will draw the following conclusion.

- (1) Both the results of the simulation experiment and the cognition model inference about braking operation show a consistent consequence. Thus, it demonstrates that this cognition model is rational to represent the driving performance at curves. Furthermore, the process of the driver’s managing information may be explained by the cognition theory and the characters or working parameters may be acquired by these kinds of simulator experiment.

TABLE 3: The results of ANOVA analysis.

ANOVA		FPRA-R30		FPPB-R30		FPRA-R50		FPPB-R50	
	Lap 1	$F(1, 10)$	P value						
1#	Group A & B	28.09	0.00**	10.5	0.01**	6.33	0.03**	6.38	0.03**
2	Group A & C	3.57	0.09*	0.99	0.34	6.63	0.03**	6.95	0.03**
3	Group B & C	0.01	0.91	1.66	0.23	0.17	0.69	0.31	0.59
Lap 2-5		$F(1, 46)$	P value						
4#	Group A & B	13.94	0.00**	34.51	0.00**	10.356	0.00**	10.58	0.00**
5	Group A & C	5.18	0.03**	11.24	0.00**	0.42	0.52	0.81	0.37
6	Group B & C	1.08	0.31	3.83	0.06*	9.35	0.00**	8.78	0.01**
Lap 1, Lap 2-5		$F(1, 28)$	P value						
7#	A:1 & A:2-5	4.86	0.04**	4.86	0.04**	8.23	0.01**	7.06	0.01**
8	B:1 & B:2-5	1.00	0.33	2.62	0.12	6.51	0.02*	2.02	0.17
9	C:1 & C:2-5	0.07	0.79	4.13	0.05**	0.41	0.53	0.92	0.35
10#	A:1 & B:2-5	28.03	0.00**	33.45	0.00**	22.23	0.00**	7.06	0.01**
11#	A:1 & C:2-5	11.77	0.00**	14.62	0.00**	18.97	0.00**	22.40	0.00**

** means P value ≤ 0.05 ; * $0.05 < P$ value ≤ 0.1 .

TABLE 4: The speed of two curves.

	Group A		Group B		Group C	
	R30	R50	R30	R50	R30	R50
Lap 1	11.06	13.03	9.21	11.08	10.97	11.83
Lap 2-5	9.66	11.39	10.20	12.37	10.05	11.85

- (2) The advance warning information can affect the driving decision and make curve driving safer by increasing the confidence value of the curve information. Known the exact process of driving performance, we can modify the driving behavior by changing the information the driver acquired.
- (3) There is more than one information channel. Drivers can acquire information via different channels instantaneously. The different information channels may have their own confidence value, and that confidence value may not be constant. Additionally, the information for the perception channel can be stored in knowledge library, or the confidence value of information may be changed by learning system.

6. Discussion

Although this pilot study has verified the cognition model in modeling driving performance and explained the information process of drivers, it still has a limitation in clarifying all details of the cognition model and the results of simulation experiment. In this simulation experiment, the impact of driver's ability had been considered and controlled in many ways. However, the personal distinction of different drivers cannot be totally controlled, and the number of the sample size still remains small. Those are the chief causes for deviation and inconsistent result of the ANOVA analysis, seen in Table 3.

We can see that the rows of all the 4 group of data, which have significant differences are 1, 4, 7, 10, and 11 with a marker of "#", the P value of which is ≤ 0.05 , and the difference of the confidence between those pairs is large enough. Others do not show a consistent result. The inconsistent may be caused by two kinds of variation coming from drivers. One is that all participants do not have the same ability; however, the analysis assumes that other modules are same during the simplification. The other is that a driver cannot keep the same state in the whole experiment; he/she may take some unpredicted mistakes. Besides, the different curvatures may affect the confidence value of the curve information.

Table 4 shows the entry speed of the curve. Both the two curves have the same result. In lap 1, group A may have the highest entry speed and group B has the lowest entry speed. However, in lap 2-5 the result is contrary. The reason may be the impact from not only the decision module but also the learning system. The analysis of this study only focuses on the braking operation by the decision module at a point time. The entry speed is the result of a sequence of operation about brake and accelerator pedals. Therefore, this involved a series of decisions based on the information. Moreover, it needs further study about other modules of the cognition models and more simulation experiments to acquire the modules working parameters.

Acknowledgment

This study is supported by the NNSFC project, the study of the mechanism for traffic signs influence on driving behavior and its cognitive model, 51108011.

References

- [1] S. G. Charlton, "The role of attention in horizontal curves: a comparison of advance warning, delineation, and road marking treatments," *Accident Analysis & Prevention*, vol. 39, no. 5, pp. 873-885, 2007.

- [2] L. Herrstedt and P. Greibe, "Safer signing and marking of horizontal curves on rural roads," *Traffic Engineering and Control*, vol. 42, no. 3, pp. 82–87, 2001.
- [3] J. E. Hummer, W. Rasdorf, D. J. Findley, C. V. Zegeer, and C. A. Sundstrom, "Curve collisions: road and collision characteristics and countermeasures," *Journal of Transportation Safety and Security*, vol. 2, no. 3, pp. 203–220, 2010.
- [4] T. Ben-Bassat and D. Shinar, "Ergonomic guidelines for traffic sign design increase sign comprehension," *Human Factors*, vol. 48, no. 1, pp. 182–195, 2006.
- [5] W. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, China Communications Press, Beijing, China, 2012.
- [6] D. D. Salvucci, "Modeling driver behavior in a cognitive architecture," *Human Factors*, vol. 48, no. 2, pp. 362–380, 2006.
- [7] D. J. Finch, P. Kompfner, C. R. Lockwood, and G. Maycock, "Speed, speed limits and accidents," Project Report 58, Transport Research Laboratory, Crowthorne, UK, 1994.
- [8] 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.
- [9] S. L. Comte and A. H. Jamson, "Traditional and innovative speed-reducing measures for curves: an investigation of driver behaviour using a driving simulator," *Safety Science*, vol. 36, no. 3, pp. 137–150, 2000.
- [10] I. R. Johnston, "Modifying driver behaviour on rural road curves: a review of recent research," in *Proceedings of the 11th Australian Road Research Board (ARRB) Conference*, vol. 11, part 4, pp. 115–124, 1982.
- [11] D. J. Findley, J. E. Hummer, W. Rasdorf, C. V. Zegeer, and T. J. Fowler, "Modeling the impact of spatial relationships on horizontal curve safety," *Accident Analysis & Prevention*, vol. 45, pp. 296–304, 2012.
- [12] J. R. Anderson, M. Matessa, and C. Lebiere, "ACT-R: a theory of higher level cognition and its relation to visual attention," *Human-Computer Interaction*, vol. 12, no. 4, pp. 439–462, 1997.
- [13] Y. C. Liu, "A simulated study on the effects of information volume on traffic signs, viewing strategies and sign familiarity upon driver's visual search performance," *International Journal of Industrial Ergonomics*, vol. 35, no. 12, pp. 1147–1158, 2005.
- [14] S. H. Chang, C. Y. Lin, C. P. Fung, J. R. Hwang, and J. L. Doong, "Driving performance assessment: effects of traffic accident location and alarm content," *Accident Analysis and Prevention*, vol. 40, no. 5, pp. 1637–1643, 2008.
- [15] M. Maltz and D. Shinar, "Imperfect in-vehicle collision avoidance warning systems can aid distracted drivers," *Transportation Research Part F*, vol. 10, no. 4, pp. 345–357, 2007.
- [16] G. Abe and J. Richardson, "The influence of alarm timing on braking response and driver trust in low speed driving," *Safety Science*, vol. 43, no. 9, pp. 639–654, 2005.
- [17] J. Roca, C. Castro, M. Bueno, and S. Moreno-Ríos, "A driving-emulation task to study the integration of goals with obligatory and prohibitory traffic signs," *Applied Ergonomics*, vol. 43, no. 1, pp. 81–88, 2012.
- [18] A. Newell, *Unified Theories of Cognition*, Harvard University Press, Cambridge, Mass, USA, 1990.
- [19] W. D. Gray, R. M. Young, and S. S. Kirschenbaum, "Introduction to this special issue on cognitive architectures and human-computer interaction," *Human-Computer Interaction*, vol. 12, no. 4, pp. 301–309, 1997.
- [20] J. R. Anderson and C. D. Schunn, "Implications of the ACT-R learning theory: no magic bullets," in *Advances in Instructional Psychology*, R. Glaser, Ed., Erlbaum, Mahwah, NJ, USA, 2000.
- [21] A. J. Hornof and D. E. Kieras, "Cognitive modeling demonstrates how people use anticipated location knowledge of menu items," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '99)*, pp. 410–417, ACM, Pittsburgh, Pa, USA, May 1999.
- [22] D. E. Kieras and D. E. Meyer, "The role of cognitive task analysis in the application of predictive models of human performance," in *Cognitive Task Analysis*, J. M. Schraagen and S. F. Chipman, Eds., pp. 237–260, Erlbaum, Mahwah, NJ, USA, 2000.

Research Article

Modeling Lane-Keeping Behavior of Bicyclists Using Survival Analysis Approach

Hongwei Guo, Wuhong Wang, Weiwei Guo, and Facheng Zhao

Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

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

Received 27 September 2012; Accepted 9 January 2013

Academic Editor: Geert Wets

Copyright © 2013 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.

Bicyclists may cross the bicycle lane and occupy the adjacent motor lanes for some reason. The mixed traffic consisting of cars and bicycles shows very complicated dynamic patterns and higher accident risk. To investigate the reason behind such phenomenon, the lifetime analysis method is adopted to examine the observed data for the behavior that bicycles cross the bicycle lane and occupy the adjacent motor lanes. The concepts named valid volume and probability of lane-keeping behavior are introduced to evaluate the influence of various external factors such as lane width and curb parking, and a semiparametric method is used to estimate the model with censored data. Six variables are used to accommodate the effects of traffic conditions. After the model estimation, the effects of the selected variables on the lane-keeping behavior are discussed. The results are expected to give a better understanding of the bicyclist behavior.

1. Introduction

Our daily life and work are closely related to traffic and mobility. Nowadays, in consequence of dramatically increasing traffic demand, traffic congestion has an immense negative impact on daily life and modern society [1, 2]. In many developing countries, like China, a typical traffic phenomenon is called “mixed traffic” which consists of cars (representing motorized vehicles) and bicycles (representing nonmotorized vehicles). Such inhomogeneous traffic flow has been considered as an important cause of traffic congestions and accidents. Additionally, bicycles are usually used as a kind of green traffic tool and it is important to improve traffic condition. Therefore, researchers turned their attention to the traffic characteristic of bicycles and mixed traffic composed of bicycles and cars. They did research from theoretical and practical points of view. Of the research on bicycles and mixed traffic, traffic simulation is popular method. Jiang et al. proposed a stochastic multivalued cellular automata model for bicycle flow [3]. Faghri and Egyházióvá developed a microscopic simulation model of mixed motor vehicle and bicycle traffic over an entire urban network [4]. Zhao et al. described mixed traffic flow by combining the NaSch model and the Burger cellular automata (BCA) model and investigated the

mixed traffic system near a bus stop [5, 6]. Mallikarjuna and Rao extended the cellular automata (CA) model to study the heterogeneous traffic [7]. On the other hand, theoretical models derived from empirical data were proposed. Oketch proposed a microscopic model to describe the mixed traffic flow by using the combination of car-following model and lateral movement [8]. Yang et al. presented a road capacity model for mixed traffic flow at the curbside stop based on queuing theory and gap acceptance theory [9]. Guo et al. used PLM model and Weibull’s distribution to analyze the lane-crossing behavior of nonmotorized vehicles under the influence of curb parking [10].

In an urban street without segregated facility, the bicyclists may drive in the motor lane because of the blockage in the bicycle lane. Once the bicyclists do not satisfy the traffic condition, they will arbitrarily change travel route and even occupy the motor lane. Particularly for the position near bus stop or parking area, the occupancy of motor lane has a strong impact on traffic performance and safety [9]. The above-mentioned literature mainly focused on analyzing the influence of mixed traffic flow on traffic performance, and most research adopted the microsimulation method. However, the research on the traffic behavior of bicyclist is

TABLE 1: Analogy between lifetime data analysis and bicyclist behavior analysis.

	Lifetime analysis	Lane-crossing behavior analysis
Parameter	Time t	bicycle volume q
Failure event	Death at time t	Lane-crossing behavior at volume q
Variable	Lifetime T	Valid volume q_c
Censoring	Lifetime T is longer than the duration of the observation	Valid volume q_c is greater than travel demand
Survival function	$S(t) = 1 - F(t)$	$F_{lk}(q)$
Probability distribution function	$F(t)$	$F_{lc}(q)$

limited. In this paper, the traffic behavior that bicyclists keep in bicycle lane (called by lane-keeping behavior) is considered. The question why bicyclists cross the bicycle lane is discussed. With this aim, a survival-analysis-based approach is used to model the lane-keeping behavior of bicyclists under various external factors. To give a quantitative analysis of the lane-keeping behavior, the probability of lane-keeping behavior is studied using field data. A concept named valid volume is proposed and a semiparametric method is used to perform the analysis. It is hoped to provide reference frame for evaluating the influence of traffic conditions on the traffic behavior of bicyclist.

2. Method

2.1. Analysis of the Bicyclist Behavior. Some bicyclists are apt to travel in the adjacent motor lanes in order to get their desired driving conditions, for example, speed and space. When the bicycle lane is blocked due to some reason, the probability of the lane-crossing behavior would increase obviously. Assume that the lane-crossing behavior occurs if the bicycle volume q is higher than a critical value q_c , yielding

$$F_{lc}(q) = \text{Pro}(q \geq q_c), \quad (1)$$

where $F_{lc}(q)$ is the distribution function of the lane-crossing behavior.

In this paper, the critical volume q_c is defined as valid volume in order to represent the maximum volume that bicycles would travel in the bicycle lane. That is, the lane-crossing behavior would not occur if the bicycle volume is lower than the valid volume ($q < q_c$). Here, another definition that the probability of lane-keeping behavior is used:

$$F_{lk}(q) = 1 - F_{lc}(q) = \text{Pro}(q < q_c). \quad (2)$$

In terms of $F_{lk}(q)$, the influence of external factors on the behavior of bicyclist can be reflected by the distribution function of a lane-crossing behavior or lane-keeping behavior. Therefore, the influential factors (e.g., narrowing lane and obstructing lane) can be analyzed from a macroscopic perspective. Considering the volume as the input variable, the data acquisition and the data processing can be simplified because some variables involving the traffic behavior of a bicyclist are difficult to be quantized.

2.2. Modeling Bicyclist Behavior Based on Lifetime Analysis. Survival analysis models (also called lifetime analysis) have been used extensively for several decades in biometrics and industrial engineering as a means of determining causality in lifetime data [11, 12]. In recent years, they have been applied in the field of transportation [13, 14], including the analysis of activity participation and scheduling, vehicle transactions analysis, and incident-duration analysis. These models concern the distribution of lifetime T :

$$F(t) = 1 - S(t) = \text{Pr}(t \geq T), \quad (3)$$

where $F(t)$ is the distribution function of lifetime data representing the probability that an individual fails before t ; $S(t)$ is the survival function representing the probability that an individual survives longer than t . $S(t)$ is also called reliability function.

The traffic behavior that bicyclists travel in bicycle lanes can be considered as a valid state under particular conditions (e.g., lane widths, traffic volume, and curb parking). Such a valid state continues with an increasing bicycle volume. If the volume is greater than the valid volume, the lane-crossing behavior will occur. It means that the particular conditions are hard to satisfy the travel demands of bicyclists. The continual process of valid state is similar to the continued life. If the lane-crossing behavior is regarded as the termination of life, the methods for lifetime data analysis can be used to estimate the valid volume q_c , which is the analogy of the lifetime T . In addition, due to the randomness of traffic behavior and the influence of curb parking, the same volume may correspond with two contrary behaviors, that is, crossing the lane or not. In this case, the lifetime analysis models are appropriate to solve such problem by censor analysis though the general statistical methods are no longer applicable [10]. The whole analogy between the bicyclist behavior analysis and lifetime data analysis is given in Table 1.

Firstly, an important concept, hazard function, is introduced. A hazard function at specified volume q in mathematical definition is

$$h(q) = \lim_{\Delta q \rightarrow 0} \frac{\text{Pro}(q \leq q_c \leq q + \Delta q \mid q_c \geq q)}{\Delta q}. \quad (4)$$

The result in the hazard function is hazard rate (or hazard), which is the instantaneous probability that the lane-crossing will occur in an infinitesimally small volume Δq after

TABLE 2: External factors and explanation.

Variable	Name	Type	Explanation
X_1	Effective width	Continuous variable	The effective width of bicycle lane
X_2	Travel speed	Continuous variable	The average travel speed for the survey interval
X_3	Car volume	Continuous variable	Car volume in the adjacent lane (veh/30 s) ^a
X_4	Curb parking	Binary indicator	1 if there are curb parking cars along the bicycle lane, 0 otherwise
X_5	Retrograde motion	Binary indicator	1 if there are retrograde bicycles in the lane at the observed interval, 0 otherwise
X_6	Safe gap	Binary indicator	1 if the adjacent lane is clear of moving car, 0 otherwise

^a“veh” is the abbreviation of vehicle.

time q . $h(q)\Delta q$ is the approximate probability of the lane-crossing behavior in $[q, q + \Delta q)$.

According to the mathematical relation between the hazard function and survival function, the probability of lane-keeping can be obtained:

$$F_{\text{lk}}(q) = \exp \left[- \int_0^q h(s) ds \right]. \quad (5)$$

To accommodate the effects of external factors, the hazard function can be written as

$$h(q) = h_0(q) g(\mathbf{x}, \boldsymbol{\beta}), \quad (6)$$

where $h_0(q)$ is the baseline hazard function, $g(\cdot)$ is a known function to represent the effects of covariates, \mathbf{x} is a column vector of covariates and it is independent of duration time, $\boldsymbol{\beta}$ is a row vector of unknown parameters. The form of (4) is one of the popular mathematical models used for duration analysis and its name is proportional hazard (PH) model.

In this study, a framework of nonparametric baseline hazard, which was proposed by Cox using $g(\mathbf{x}, \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}\mathbf{x})$, is adopted [15]. With this parameterization, the hazard function is

$$h(q) = h_0(q) \exp(\boldsymbol{\beta}\mathbf{x}). \quad (7)$$

The endurance probability function combining (5) and (7) can be written as

$$S(q) = [S_0(q)]^{\exp(\boldsymbol{\beta}\mathbf{x})}, \quad (8)$$

where $S_0(q)$ is the baseline probability function for the lane-keeping behavior. It represents the probability without any external influence.

The shape of $h_0(q)$ in the PH model has important implications for data analysis. Also a parametric shape could be chosen according to data distribution. In this paper, the nonparametric baseline hazard is used to avoid the error when the assumed parametric form is incorrect. The parameter estimation can use the partial likelihood method. Other methods can be referred to in [11, 12].

2.3. External Factors Selection. The selection of external factors takes into account the previous researches and arguments regarding the effects of the exogenous variables and human

factors on bicyclist behavior. Three broad sets of variables may influence the bicyclist behavior: personal characteristics, traffic conditions, and trip characteristics. In this paper, the traffic conditions are considered. The following factors, as shown in Table 2, are adopted to construct the model.

3. Survey and Data

The field survey is conducted in the urban roads with no isolation facilities. The selected survey sites are monitored by video cameras. Then, the bicycle volumes in the lanes with different effective widths can be acquired. According to [10], the effective width of bicycle lane is defined as the physical width minus the margin of safety (0.5 m). Such safety margin indicates the influence of curb parking so that bicycles would keep a safe distance from the parked cars. On the other hand, the data related to the external factors are also derived from the video survey. The data acquisition is performed by manual counting and recording and the assistance of video processing tool.

The length of the observed section is 25 m and there is no influence of bus stop and pedestrian crosswalk. In consideration of the discrete arrival of bicycles and the nonuniform volume, short observed interval may not include enough samples while long interval may influence the definition of data status. Therefore, the observed interval is 30 s. The status of each interval is defined as (a) censored data if there is no bicycle entering the motor lane in the interval and (b) distinct data if the lane-crossing behavior occurs in the interval [10]. The field surveys are conducted in four typical urban roads in Beijing, including East Jiaoda Road, North Yufang Road, West Tucheng Road, and Da Liushu Road. The basic features of observed sections are shown in Table 3.

4. Model Estimation and Discussion

4.1. Model Estimation. Table 4 shows the model estimation of the lane-keeping behavior of bicyclists. The LR statistic of the estimated model clearly indicates the overall goodness-of-fit (the LR statistic is 98.4, which is greater than the chi-squared statistic with 6 degrees of freedom at any reasonable level of significance). The statistical significance of each variable is examined by the Z test, which has an asymptotically normal distribution with mean zero and variance one. The significant

TABLE 3: Basic features of observed sections.

Effective width (m)	Sample size	Volume distribution (veh/30 s)			Lane-crossing ratio (%)
		Min	Max	Mean	
1.5	118	2	28	12	47
2.5	378	4	35	17	33
3	381	2	34	26	28
3.5	430	3	39	19	29

TABLE 4: Model estimation.

Variable	Estimate	Standard error	Z test	P value
X_1	-0.271	0.12	-2.264	0.024
X_2	0.002	0.152	2.092	0.036
X_3	-0.201	0.192	1.047	0.295
X_4	0.317	0.001	-2.238	0.025
X_5	0.414	0.001	-3.777	<0.001
X_6	0.148	0.148	-1.014	0.303

level corresponding to each covariate is given by P value in Table 4. From the results, most of the included variables are statistically significant at the 0.05 level of significance. It means that these covariates are significantly related to the violation behavior. Two covariates (X_3 car volume and X_6 safe gap) have a relatively low significant level. The effects of the variables on lane-crossing behavior will be discussed in the next section.

Figure 1 shows the lane-keeping probability by the proposed model with the average of all variables. It means that the curve in Figure 1 can reflect the average lane-keeping probability of the typical bicycle flow which has an average value for every external factor. The curve is monotonely decreasing. The median of the distribution is 25 vehicle/30 s, indicating that over a half of the observed interval would result in lane-crossing behavior if the bicycle volume is greater than 25 veh/30 s in average condition (2.5 m lane, bicycle speed is 12 km/h, car volume is 450 pcu/h/lane, and about 30% of the bicyclists are affected by curbside parking). If the variable changes (representing the change of external factor), the probability of lane keeping would change correspondingly.

4.2. Effects of External Factors. In the proportional hazard model, the effects of variables are multiplicative on the baseline hazard function. A negative coefficient on a variable implies that an increase in the corresponding variables decreases the hazard rate, or equivalently increases the valid volume. The greater valid volume means that the occurrence of lane-crossing behavior decreases. The effects of external factors are analyzed in the following.

(1) Effective width of bicycle lane shows a significant negative on hazard. It means the wider the lane width for bicycles, the less the lane-crossing behavior would occur. Figure 2 shows various distributions of the lane-keeping behavior with different lane widths. According to the empirical data and the estimated model, if the effective width of a bicycle

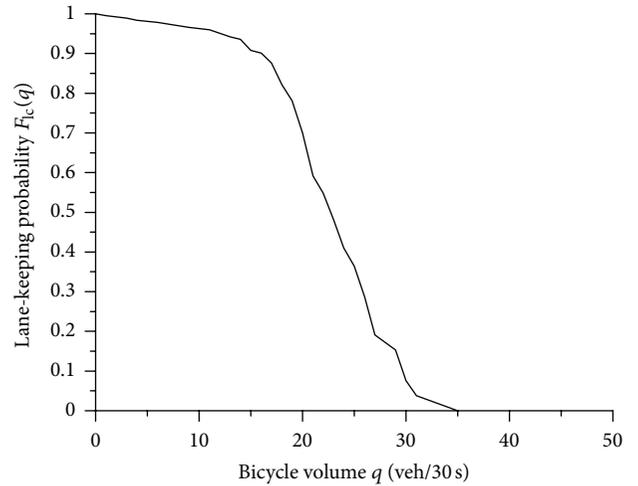


FIGURE 1: Distribution of the lane-keeping probability in average condition.

lane decreases from 3.5 m to 2.0 m, the valid volume will decline from 28 veh/30 s to 20 veh/30 s (the valid probability is 0.8). Assume that the average volume is 15 veh/30 s, the probability that all bicycles travel in the bicycle lane would be 0.6 under the condition of the narrow lane. Meanwhile, the valid volume would decrease by 40%.

(2) Curbside parking also shows a positive effect on the hazard while it means that the curbside parking can increase the hazard or decrease the probability of lane-keeping behavior. The effect of curbside parking on the probability of the lane-keeping behavior is shown in Figure 3. The road sections for a comparative analysis show differences in the distribution of valid volume. Namely, probabilities of the lane-keeping behavior in the sections with curbside parking are lower than those without influence of curbside parking. It should be noted that the probability that the lane-crossing behavior occurs is low when the volume of the bicycle is low. Meanwhile, the influence of curbside parking is insignificant. When the volume of the bicycle is high, the relation between the occurrence of the lane-crossing behavior and curbside parking is also insignificant. From the results, the influence of curbside parking on the valid probability is significant when the volume of bicycle distributes in median range. Taking the width of 3.0 m as an example, the influence of curbside parking on the lane-keeping behavior is significant when the volume ranges between 22 and 32 veh/30 s.

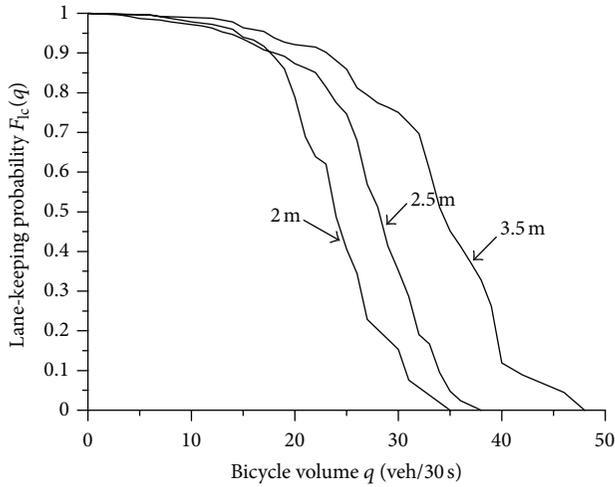


FIGURE 2: Distributions of the lane-keeping probability with various lane widths.

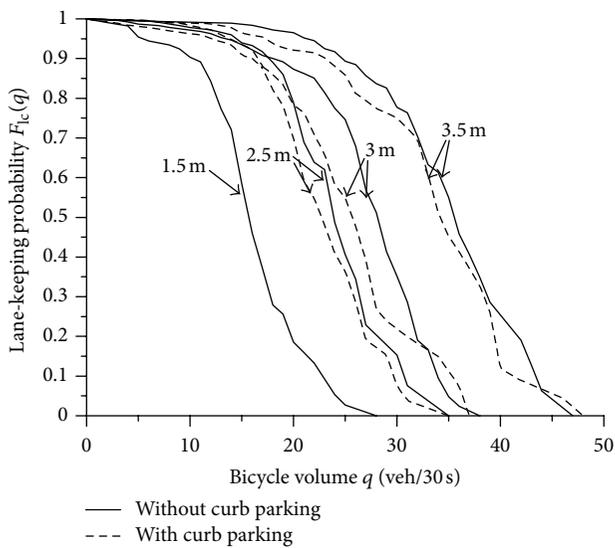


FIGURE 3: Distributions of lane-keeping probability under influence of curb parking.

(3) The travel speed of a bicycle can have a positive effect on the hazard function. The faster the bicycles travel, the higher the possibility they cross the bicycle lane. In this paper, the electric bicycle is considered as bicycle. According to the field survey, there is a certain number of electric bicycles that cross the bicycle lane and travel in the motor lane. The electric bicycles travel faster than other bicycles; thus, they want to seek ideal travel space. Particularly, when there are bicycles travelling in low speed or curb parking car in front of the faster electric bicycles, they change the travel direction and overtake the blockage via the motor lane without the least hesitation.

(4) Retrograde motion has a positive effect on the hazard function, like travel speed of bicycle. As shown in Figure 4,

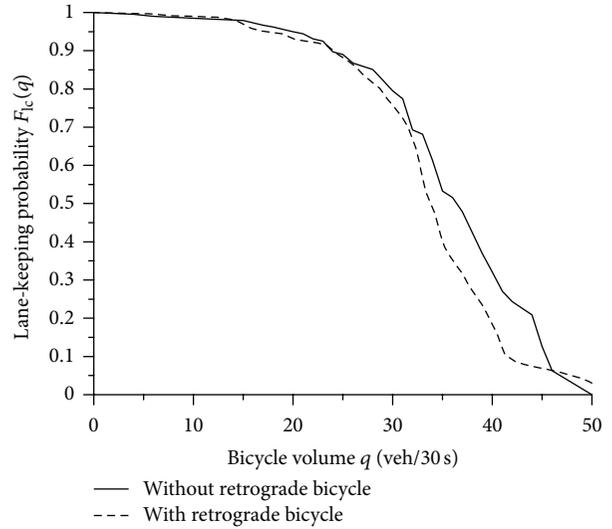


FIGURE 4: Effect of retrograde motion on lane-keeping behavior.

the existence of retrograde bicycle can decrease the probability of lane-keeping behavior. Such effect is more significant in the condition of higher bicycle volume. The retrograde motion of bicycles can hinder the travel routes of other bicycles; it is easy to provide a motivation for changing travel route, even changing the lane.

(5) According to the estimation results, the effect of car volume and safe gap is not significant from the perspective of statistic. However, the car volume still can influence the bicyclist behavior. If the car volume is high, the bicyclist may have little chance to travel in the motor lane. Additionally, the variable of safe gap can also reflect the chance and the safety for a bicyclist to travel in the motor lane. From the field survey and the estimated results, a certain number of bicycles travel in the motor lane when the car volume is very high. These bicyclists neglect the accident risk caused by the lane-crossing behavior. It is dangerous for the bicyclists to travel in the motor lane in heavy car flow. And the lane-crossing behavior can enforce a blockade against the moving car so that the traffic performance reduces obviously.

5. Conclusion

This paper proposed a model to describe the lane-keeping behavior of a bicyclist in urban street by using survival analysis. A concept of valid volume is also proposed to describe the relation between the lane-crossing behavior and the bicycle volume. The volume data are defined as censored data and uncensored data. Proportional hazard method is used to estimate the field data with censored data. In order to capture the effect of external factors involving traffic conditions, six variables are selected to construct the PH model. The results show that the effective width of bicycle lane, travel speed, curbs parking, and retrograde motion have significant effect on the lane-keeping behavior. Two variables (car volume and safe gape) show relatively low significance. It is concluded that the lane-keeping behavior results from

various related factors such as personal features, traffic conditions, and environmental factors, and any change of the influential factors can modify the lane-keeping behavior. Therefore, the planning and designing of urban street should consider these influential factors apprehensively.

The future work will focus on the influential factors. More factors will be introduced into the model and the field surveys of sites will be increased to obtain more empirical data. For example, the average speed of bicycle travelling in the car lane could be an important influential factor on the lane-keeping behavior of cyclists. Also the significances of variables and their effects on bicyclist behavior will be discussed deeply.

Acknowledgments

The authors would like to thank the anonymous editor and referees for their valuable comments. This research is supported by the Programme of Introducing Talents of Discipline to Universities under Grant no. B12022 and the Programme of International Science and Technology Cooperation in the Beijing Institute of Technology.

References

- [1] W. H. Wang, W. Zhang, H. W. Guo, H. Bubb, and K. Ikeuchi, "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [2] 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.
- [3] R. Jiang, B. Jia, and Q. S. Wu, "Stochastic multi-value cellular automata models for bicycle flow," *Journal of Physics A*, vol. 37, no. 6, pp. 2063–2072, 2004.
- [4] A. Faghri and E. Egyházióvá, "Development of a computer simulation model of mixed motor vehicle and bicycle traffic on an urban road network," *Transportation Research Record*, no. 1674, pp. 86–93, 1999.
- [5] X. M. Zhao, B. Jia, Z. Y. Gao, and R. Jiang, "Traffic interactions between motorized vehicles and nonmotorized vehicles near a bus stop," *Journal of Transportation Engineering*, vol. 135, no. 11, pp. 894–906, 2009.
- [6] X. M. Zhao, B. Jia, Z. Y. Gao, and R. Jiang, "Congestions and spatiotemporal patterns in a cellular automaton model for mixed traffic flow," in *Proceedings of the 4th International Conference on Natural Computation (ICNC '08)*, pp. 425–429, October 2008.
- [7] C. Mallikarjuna and K. R. Rao, "Cellular Automata model for heterogeneous traffic," *Journal of Advanced Transportation*, vol. 43, no. 3, pp. 321–345, 2009.
- [8] T. G. Oketch, "New modeling approach for mixed-traffic streams with nonmotorized vehicles," *Transportation Research Record*, no. 1705, pp. 61–69, 2000.
- [9] X. Yang, Z. Gao, X. Zhao, and B. Si, "Road capacity at bus stops with mixed traffic flow in China," *Transportation Research Record*, no. 2111, pp. 18–23, 2009.
- [10] H. W. Guo, Z. Y. Gao, X. M. Zhao, and X. B. Yang, "Traffic behavior analysis of non-motorized vehicle under influence of curb parking," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 1, pp. 79–84, 2011.
- [11] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New Jersey, NJ, USA, 2002.
- [12] C. R. Bhat, "Duration modeling," in *Handbook of Transport Modelling*, K. J. Button and D. A. Hensher, Eds., pp. 91–111, Elsevier Science, Amsterdam, The Netherlands, 2000.
- [13] D. A. Hensher and F. L. Mannering, "Hazard-based duration models and their application to transport analysis," *Transport Reviews*, vol. 14, no. 1, pp. 63–82, 1994.
- [14] D. Nam and F. Mannering, "An exploratory hazard-based analysis of highway incident duration," *Transportation Research Part A*, vol. 34, no. 2, pp. 85–102, 2000.
- [15] D. R. Cox, "Regression Models and Life Tables," *Journal of the Royal Statistical Society B*, vol. 34, no. 2, pp. 187–220, 1972.

Research Article

Individual Subjective Initiative Merge Model Based on Cellular Automaton

Yin-Jie Xu,¹ Yu-Guang Chen,¹ Chao Yang,² and Yi-Chao Pu²

¹ Pohl Institute of Solid State Physics, Tongji University, Shanghai 200092, China

² Key Laboratory of Road and Traffic Engineering, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Chao Yang; tongjiyc@163.com

Received 2 August 2012; Revised 2 January 2013; Accepted 3 January 2013

Academic Editor: Wuhong Wang

Copyright © 2013 Yin-Jie Xu 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 merge control models proposed for work zones are classified into two types (Hard Control Merge (HCM) model and Soft Control Merge (SCM) model) according to their own control intensity and are compared with a new model, called Individual Subjective Initiative Merge (ISIM) model, which is based on the linear lane-changing probability strategy in the merging area. The attention of this paper is paid to the positive impact of the individual subjective initiative for the whole traffic system. Three models (ISIM, HCM, and SCM) are established and compared with each other by two order parameters, that is, system output and average vehicle travel time. Finally, numerical results show that both ISIM and SCM perform better than HCM. Compared with SCM, the output of ISIM is 20 vehicles per hour higher under the symmetric input condition and is more stable under the asymmetric input condition. Meanwhile, the average travel time of ISIM is 2000 time steps less under the oversaturated input condition.

1. Introduction

Over the past several decades, many methods and rules [1–9] have been proposed for the work zone system which consists of a double-lane (DL) model and a single-lane (SL) model. Among them, many have been implemented or tested in practice and have been effective on the real traffic system. They are named after the characteristics of the model, the strategy, the authors and so on, but when we focus on the control intensity, we can briefly classify them into two types: (1) Hard Control Merge (HCM) model (whose control intensity is strong, such as lane-based signal system [10]); (2) Soft Control Merge (SCM) model (whose control intensity is mild, such as dynamic late merge system and lane-based dynamic merge system [11]). But both models ignore the impact of the individual subjective initiative for the whole system and it is the point that we focus on in this paper.

In next sections, three models with different lane-changing rules and a vehicle input method are introduced. Simulation results and analyses are presented after that. Conclusions and further studies are given in the last section.

2. Model

We introduce Cellular Automaton (CA) [12, 13] which is a discrete model and present three models based on CA. The CA model consists of a regular grid of cells with a finite number of states. For each cell, a set of cells called its neighborhood is defined relative to it. The selection of the initial state is determined by the model. A new generation is created according to some fixed rules which determine the new state of each cell with the consideration of the current states of the cell and its neighborhood. Typically, the rule to update the state of cells is the same for each cell and does not change over time. Above all, it is applied to the whole grid simultaneously.

2.1. Basic Updating Rules. The basic updating rules are introduced according to NS model [14], and the updating rule for the n th vehicle at $(t + 1)$ th time step is as follows.

- (1) Accelerating at the $(t + 1/4)$ th time step

$$V_{n,t+1} = V_{n,t} + 1. \quad (1)$$

(2) Safety breaking at the $(t + 2/4)$ th time step

$$V_{n,t+1} = \min(V_{n,t}, d_{n,t}). \quad (2)$$

(3) Random decelerating at the $(t + 3/4)$ th time step

$$V_{n,t+1} = \max(V_{n,t+1} - 1, 0) \quad \text{with probability } P_d \in [0, 1],$$

$$V_{n,t+1} = V_{n,t+1} \quad \text{with probability } 1 - P_d. \quad (3)$$

(4) Moving at the $(t + 1)$ th time step

$$X_{n,t+1} = X_{n,t} + V_{n,t+1}, \quad (4)$$

where $V_{n,t}$ is the velocity of the n th vehicle at t th time step, $X_{n,t}$ is the position of the n th vehicle at t th time step, $d_{n,t}$ is the forward headway of the n th vehicle at t th time step, and P_d is the decelerating probability.

2.2. Basic Model. The basic model is presented in Figure 1. The work zone system is composed of a 100-lattice long DL model and a 100-lattice long SL model (one lattice is equal to 7.5 meters). We name the upper lane as Lane 1 and the other as Lane 2. We divide DL model into two parts: Part 1 is 94-lattice long and Part 2 is 6-lattice long. Part 1 is the core area of the model, and the basic model converts into different merge models when we introduce different lane-changing rules into Part 1. We use different kinds of lane-changing probability ($P_l \in [0, 1]$) to simulate different lane-changing rules. The higher the lane-changing probability is, the more probably the vehicle changes its lane. Part 2 is the coercive lane-changing area. In this part, vehicles on Lane 1 have to change lane before the merge point and vehicles on Lane 2 are not allowed to change lane. Thus, in Part 2, the lane-changing probability of the vehicle on Lane 1 is always 1, and the probability of the vehicle on Lane 2 is 0.

2.3. Merge Models. We introduce three different kinds of lane-changing rules into the basic model and then get three models: (1) ISIM, (2) HCM, and (3) SCM.

(1) *ISIM.* The lane-changing probability of ISIM is shown in Figure 2. Then we set the probability of the vehicle on Lane 1 changing to Lane 2 as follows:

$$P_{l,1 \rightarrow 2} = 0.5 + 0.5 \times \frac{L}{94} \quad L \leq 94, \quad (5)$$

$$P_{l,1 \rightarrow 2} = 1 \quad L > 94,$$

and the probability of the vehicle on Lane 2 changing to Lane 1 is

$$P_{l,2 \rightarrow 1} = 0.5 - 0.5 \times \frac{L}{94} \quad L \leq 94, \quad (6)$$

$$P_{l,2 \rightarrow 1} = 0 \quad L > 94.$$

With the linear lane-changing probability, we try to perform a more realistic simulation and find out the impact of the individual subjective initiative on the whole system.

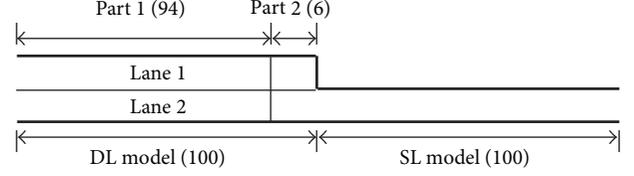


FIGURE 1: The basic work zone system.

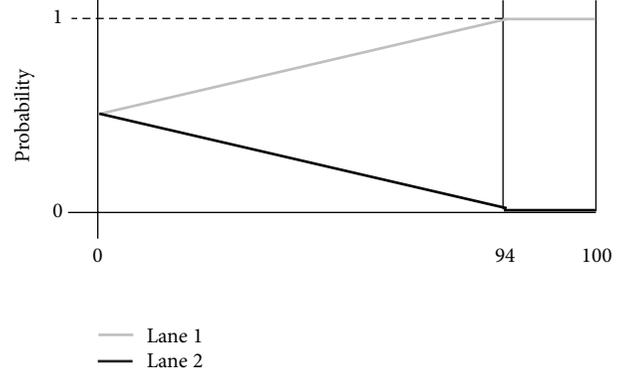


FIGURE 2: Lane-changing probability of ISIM.

(2) *HCM.* The lane-changing probability of HCM is presented in Figure 3. The two lanes are separated in Part 1. According to this, the lane change is not allowed and the vehicle will stay on the lane where it is injected at the beginning. We put a signal between the 94th lattice and the 95th lattice, and only one lane is allowed to drive in each interval according to the signal (the interval between two adjacent signal changes is set as 30 time steps). We can use this model to simulate lane-based signal system.

The probability of the vehicle on Lane 1 changing to Lane 2 is

$$P_{l,1 \rightarrow 2} = 0 \quad L \leq 94, \quad (7)$$

$$P_{l,1 \rightarrow 2} = 1 \quad L > 94,$$

and the probability of the vehicle on Lane 2 changing to Lane 1 is

$$P_{l,2 \rightarrow 1} = 0. \quad (8)$$

(3) *SCM.* The lane-changing probability of SCM is illustrated in Figure 4. The probability of the vehicle on Lane 1 changing to Lane 2 is

$$P_{l,1 \rightarrow 2} = 0.5 + 0.5 \times \frac{L}{94} \quad L \leq 94, \quad (9)$$

$$P_{l,1 \rightarrow 2} = 1 \quad L > 94,$$

and the probability of the vehicle on Lane 2 changing to Lane 1 is

$$P_{l,2 \rightarrow 1} = 0. \quad (10)$$

The difference between ISIM and SCM is the lane-changing probability of the vehicle on Lane 2. When the traffic

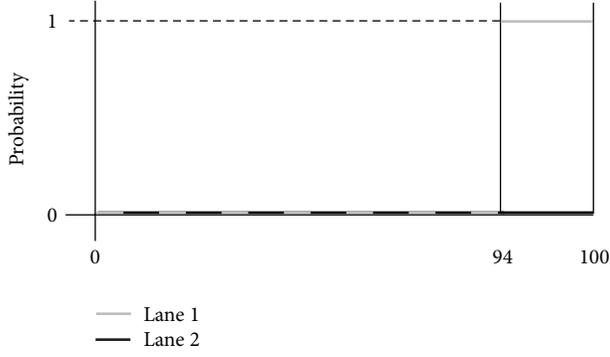


FIGURE 3: Lane-changing probability of HCM.

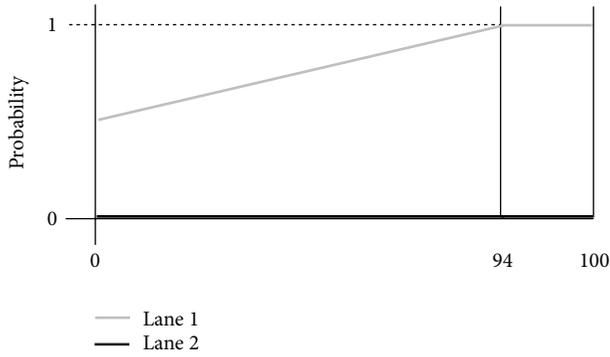


FIGURE 4: Lane-changing probability of SCM.

demand is light, vehicles on Lane 1 can easily change lane at the upstream of the DL model (like the static early merge in [11]). When the traffic demand is high and the model is congested, they may not easily change to Lane 2 at the upstream of the DL model and have to change lane near the end of the DL model (like the static late merge in [11]). Thus, with the increasing of the traffic demand, the real merge point will move backwards constantly. Meanwhile, the static early merge will convert into the static late merge gradually. According to this, we can use it to simulate the lane-based dynamic merge in some degree but we do not set a definite control threshold.

2.4. Vehicle Input Methods. We introduce the probability distribution function here to obtain the quantity of the arriving vehicle per time step. According to the traffic flow theory, the Poisson distribution function is adaptive under the light traffic condition and the binomial distribution function is chosen under the congested condition [15]. Here we set the threshold as 0.25 vehicles per time step (equal to 900 vehicles per hour (vph)).

The Poisson distribution function is

$$P(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \quad (11)$$

and the binomial distribution function is

$$P(x) = C_n^x \left(\frac{\lambda t}{n} \right)^x \left(1 - \frac{\lambda t}{n} \right)^{n-x}, \quad (12)$$

where $P(x)$ is the probability that x vehicles arrive within t time steps, λt is the quantity of the arriving vehicle within t time steps (here $\lambda = 0.25$), and n is the maximum quantity of the arriving vehicle per time step.

After getting the quantity of arriving vehicles, we do not input them into the model directly. Instead, we put them into a stack and set the serial number and arriving time for them. Then we input them into the model in turn (according to the First-In-First-Out principle) with the consideration of the real-time conditions of the lane and the stack.

3. Simulation Results and Analysis

We performed simulations based on the three models described above with $v_{\max} = 4$ lattice/time step. Then the results, such as the output and the travel time, are collected under different input conditions. All data are collected during the 3600 time steps from the 14401st time step to the 18000th time step.

3.1. Theoretical Output of Work Zone System. We simulate two other basic models for comparison: a 200-lattice long SL model and a 200-lattice long DL model. Briefly, the work zone model is composed of a SL model and a DL model, so the simulation of SL and DL model helps us to get the theoretical limitations of the output of the work zone system in Figure 5(a). For a better discussion, we make the fitting line of the output of SL and DL model in Figure 5(b). From it, we notice that $\text{Output}_{\max, \text{DL}} (= 2600 \text{ vph})$ is larger than $\text{Output}_{\max, \text{SL}} (= 1160 \text{ vph}) \times 2$. A well-performing DL model can reduce the negative impact of the following vehicle [16]. According to this, it is the logical result of taking full advantage of DL model with safety lane-changing rules. The output of work zone system will be suppressed by the SL model, so the theoretical Output_{\max} of the work zone system is similar to 1160 vph. Compared with a SL model, a DL model can support a higher single-lane input value with the reasonable merge rules, so the theoretical value of work zone may be a little more than 1160 vph. We cannot get the exact number, so we use the result of the SL model as the theoretical value of the work zone system.

3.2. Output under Symmetric Input Conditions. We simulate the three models individually, and the results are illustrated in Figure 6(a). In Figure 6(b), we compare the results of three models with the theoretical value. The $\text{Output}_{\max, \text{ISIM}}$ is the largest, and both the $\text{Output}_{\max, \text{ISIM}}$ and the $\text{Output}_{\max, \text{SCM}}$ are better than the theoretical value. The good performance of ISIM and SCM is due to the individual subjective initiative, and vehicles in these two models can change lane with safety lane-changing rules. This behavior takes full advantage of DL model and led to better output results. As illustrated in Figure 6, it seems that ISIM result is only slightly better than that of SCM. However, we will find that the advantage of ISIM is more obvious compared with SCM when we observe the travel time presented in the following section. Compared with the others, the result of HCM is unsatisfactory and it is more inefficient not only than the other two models but

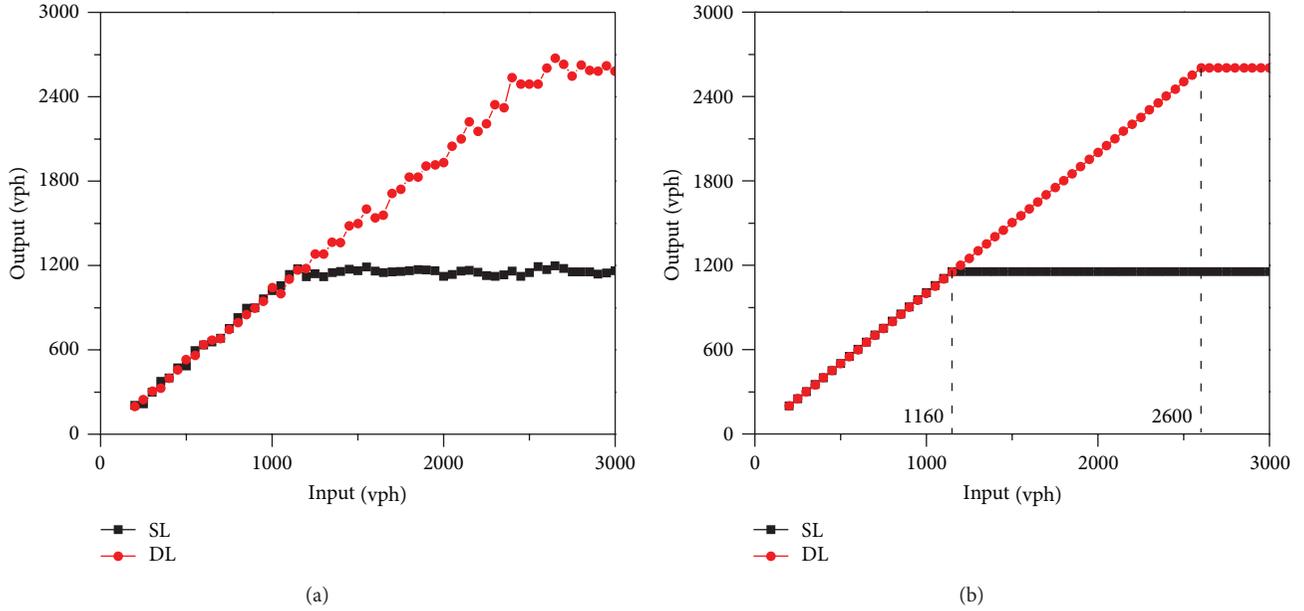


FIGURE 5: (a) Output of SL and DL model; (b) the fitting line of SL and DL model.

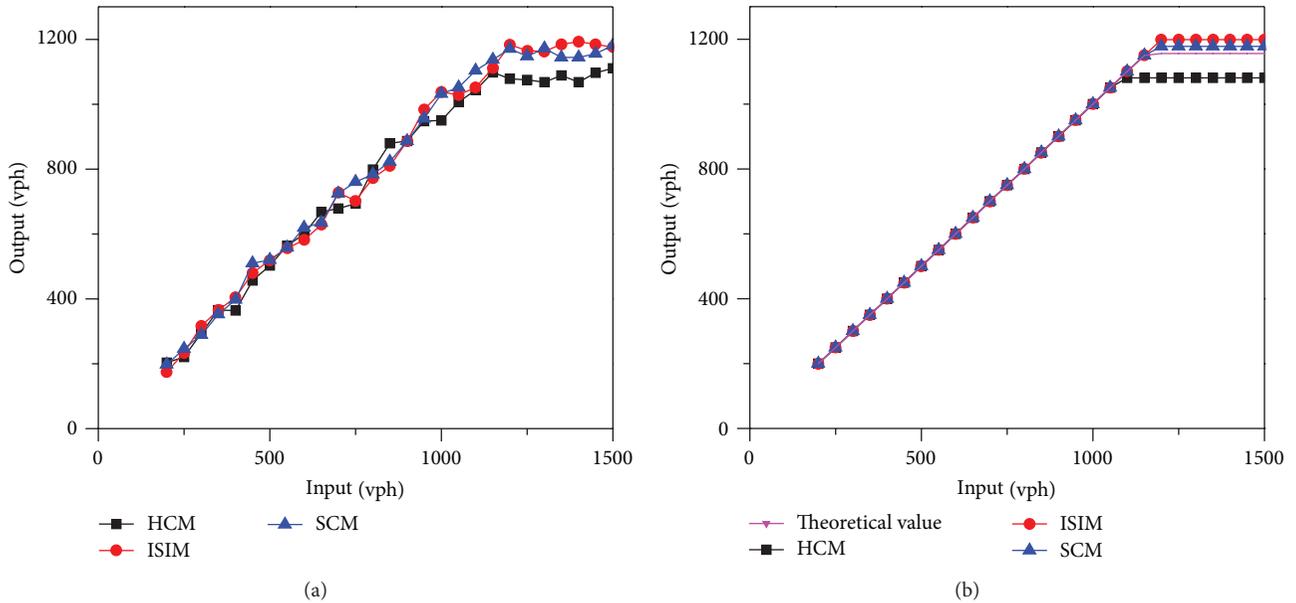


FIGURE 6: (a) Output results of three models; (b) fitting lines of three models and the theoretical value ($\text{Output}_{\max, \text{ISIM}} = 1200$ vph, $\text{Output}_{\max, \text{SCM}} = 1180$ vph, $\text{Output}_{\max, \text{HCM}} = 1080$ vph, $\text{Output}_{\max, \text{theoretical}} = 1160$ vph).

also than the theoretical value. In HCM, vehicles have to slow down or even stop to wait the “PASS” signal, and finally this behavior results in the lowest output.

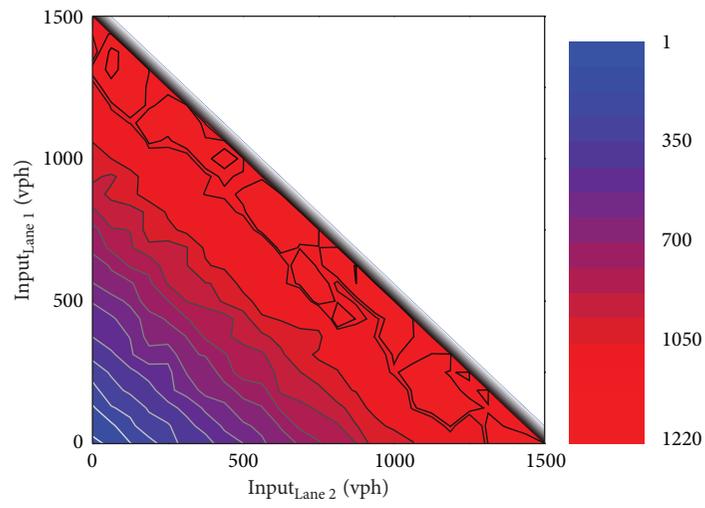
3.3. Output under Asymmetric Input Conditions. Next, we run three models under asymmetric input conditions as (13) and set the input as 1500 vph according to the theoretical value mentioned above:

$$\text{Input}_{\text{Lane 1}} + \text{Input}_{\text{Lane 2}} \leq 1500 \text{ vph.} \quad (13)$$

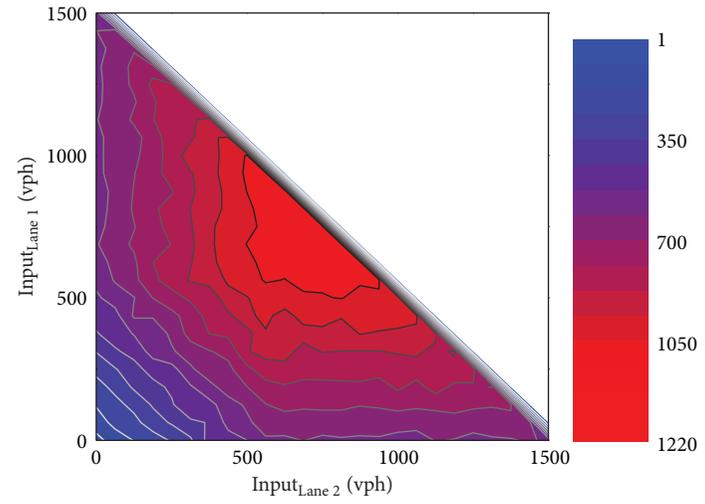
The results are presented as in Figure 7.

From Figure 7(b) we can find that HCM is sensitive to the asymmetric input. Once the inputs of two lanes are asymmetric, the output will decrease quickly. Then, we present the sectional view in Figure 8 where $\text{Input}_{\text{Lane 1}} + \text{Input}_{\text{Lane 2}} = 1500$ vph for the better observation.

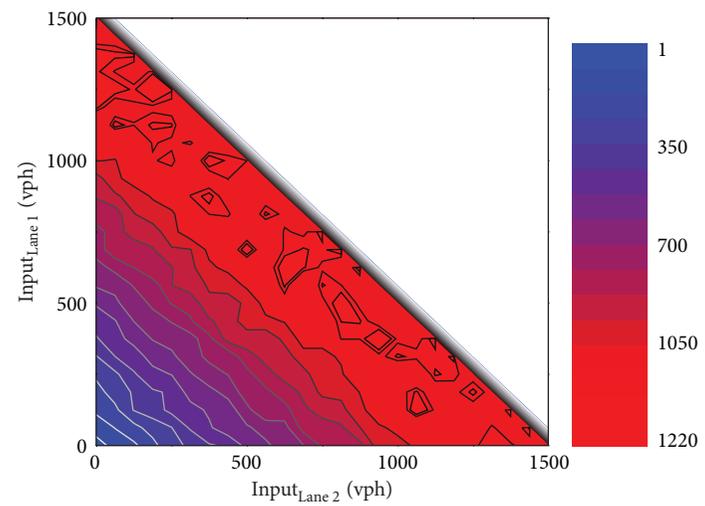
It is obvious from the Figure 8 that, for HCM, the output approaches the maximum when the inputs of each lanes are similarly equal, and it is noticed that the asymmetric input condition results in the huge fluctuation of the output. The



(a)



(b)



(c)

FIGURE 7: Output results of three models under asymmetric input conditions (a) ISIM, (b) HCM, and (c) SCM.

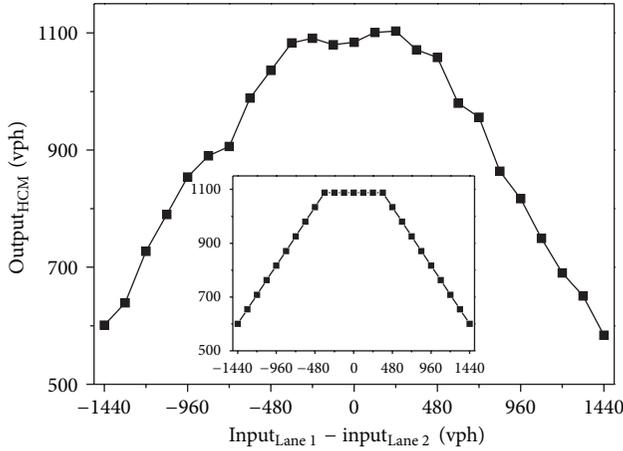


FIGURE 8: Output of HCM in the section where $\text{Input}_{\text{Lane } 1} + \text{Input}_{\text{Lane } 2} = 1500$ vph, and the insert shows its fitting line.

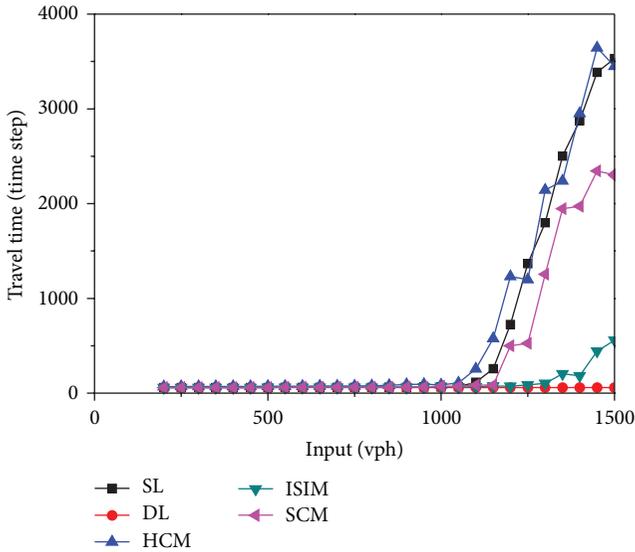


FIGURE 9: The travel time results of SL model, DL model, and three models.

maximum is bigger than the minimum by 27.3%, and the descending slope measured from the fitting line in the insert is equal to -1 approximately. According to this, we should balance the input of each lane carefully when HCM model is implemented. SCM is not sensitive to the asymmetric input as the illustration. But we still can find out that when $\text{Input}_{\text{Lane } 1} > \text{Input}_{\text{Lane } 2}$ the output of SCM is a bit better than that in the case of $\text{Input}_{\text{Lane } 1} < \text{Input}_{\text{Lane } 2}$. It is the result that the vehicle has to change lane early when more vehicles are injected into Lane 1, and this behavior makes the DL model start to take effect passively. Compared with the previous two models, ISIM is more satisfactory. It is not sensitive to the asymmetric input condition and can get a better output result even than SCM. With the more optional lane-changing rules, vehicles on each lane can balance themselves spontaneously.

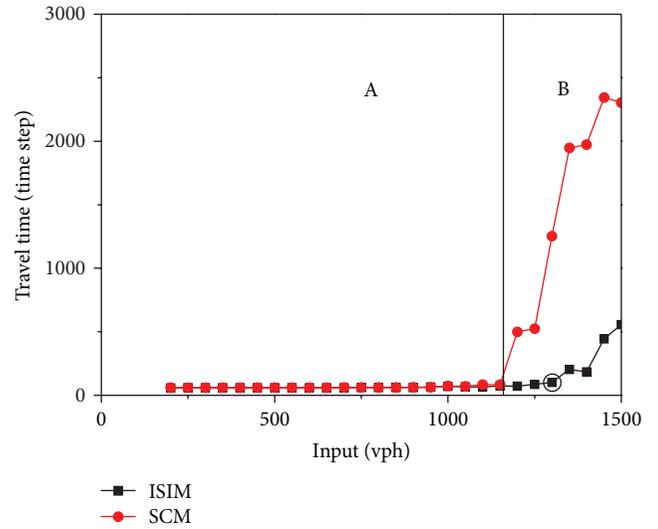


FIGURE 10: The travel time results of ISIM and SCM.

Thus the DL model works better, and the advantage will be more apparent when we focus on the travel time result.

3.4. Travel Time under Symmetric Input Conditions. Similar to the discussion of output, we simulate SL and DL model to obtain the theoretical travel time and the results of SL and DL model and other three models are presented in Figure 9.

When the input increases to their respective $\text{Output}_{\text{max}}$, the travel time begins to increase as shown in Figure 9 and the travel time of HCM increases especially quickly. It can be estimated that if the work zone model is well established, its travel time will be somewhere between the DL and SL model. Then we focus on the travel time results of three models. The results of HCM and SL model are overlapping, while the results of ISIM and SCM are optimized, and they both appear between the value of DL and SL model. According to this, we illustrate the results of ISIM and SCM in Figure 10 separately.

In Figure 10, we divide the input space into two areas according to the theoretical value of SL model:

Area A: $0 \text{ vph} < \text{Input} < 1160 \text{ vph}$;

Area B: $1160 \text{ vph} < \text{Input} < 1500 \text{ vph}$.

In Area A, the input is smaller than the theoretical value, and the travel time results of both models are about 58 time steps, so the mean velocity at which the vehicle passes through the system is about 3.4 lattice per time step. According to the maximum velocity (4 lattice per time step), it is a satisfactory value. In Area B, once the input is larger than the theoretical value, the travel time result of SCM increases dramatically. The result of ISIM is obviously better. The circled point (after which the result increases rapidly) of ISIM is about 1350 vph which is about 200 vph higher than the theoretical value and that of SCM. Furthermore, the travel time of ISIM is 2000 time steps less when the input reaches to 1500 vph, and this is a huge advantage to move vehicles in less time. It is noticed that the individual subjective initiative in ISIM can improve the velocity and capacity of the traffic.

4. Conclusions and Further Studies

In this paper, we try to make a comparison between the three models: ISIM, HCM, and SCM. HCM model performs unsatisfactorily in both output and average travel time, because its control method is purely mechanical. The vehicle in it has to slow down or even stop for the signal, and the individual subjective initiative is not considered. When HCM is implemented, many factors should be taken into consideration, such as the careful input control, and the interval between the signal changes. Meanwhile, SCM model, owing to the passive performance of DL model, performs better than HCM model when the traffic is heavy. ISIM model performs the best among the three models under different traffic conditions due to the well-designed individual subjective initiative. The huge advantage in travel time results compared with SCM cannot be ignored. Vehicles in it can balance themselves more actively, and this makes the system perform the best in both output results, and travel time results. The individual subjective initiative is helpful for the vehicle to make a good use of system resources and make the system more flexible.

In further studies, we will try to take more details of ISIM model into consideration, such as the complex driver-vehicle behavior [17]. We try to establish a more realistic model and make a reasonable reference model for further studies.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (70631002, 10774112) and the Program for New Century Excellent Talents in University (NCET-08-0406).

References

- [1] A. Tarko, S. Kanipakapatman, and J. Wasson, "Modeling and optimization of the Indiana Lane merge control system on approaches to freeway work zones," Final Report FHWA/IN/JTRP-97/12, Purdue University, West Lafayette, Ind, USA, 1998.
- [2] P. T. McCoy, G. Pesti, and P. S. Byrd, "Alternative information to alleviate work zone related delays," SPR-PL-1 SPR-PL-1(35) P513, University of Nebraska-Lincoln, 1999.
- [3] G. Pesti, D. R. Jessen, P. S. Byrd, and P. T. McCoy, "Traffic flow characteristics of the late merge work zone control strategy," *Transportation Research Record*, no. 1657, pp. 1–9, 1999.
- [4] C. H. Walters, V. J. Pezoldt, K. N. Womack, S. A. Cooner, and B. T. Kuhn, "Understanding road rage: summary of first-year project activities," Tech. Rep. TX-01/4945-1, Texas Transportation Institute, 2000.
- [5] A. Tarko and S. Venugopal, "Safety and capacity evaluation of the Indiana lane merge system," FHWA/iN/JTRP/-2000/19, Purdue University, West Lafayette, Ind, USA, 2001.
- [6] T. Datta, K. Schattler, P. Kar, and A. Guha, "Development and Evaluation of an advanced dynamic lane merge traffic control system for 3 to 2 lane transition areas in work zones," Report RC-1451, Michigan Department of Transportation, 2004.
- [7] A. G. Beacher, M. D. Fontaine, and N. J. Garber, "Field evaluation of late merge traffic control in work zones," *Transportation Research Record*, no. 1911, pp. 33–41, 2005.
- [8] A. G. Beacher, M. D. Fontaine, and N. J. Garber, "Guidelines for using late merge traffic control in work zones: results of a simulation-based study," *Transportation Research Record*, no. 1911, pp. 42–50, 2005.
- [9] S. Ishaka, Y. Qi, and P. Rayaproluc, "Safety evaluation of joint and conventional lane merge configurations for freeway work zones," *Traffic Injury Prevention*, vol. 13, no. 2, pp. 199–208, 2012.
- [10] N. Yang, G. L. Chang, and K. P. Kang, "Simulation-based study on a lane-based signal system for merge control at freeway work zones," *Journal of Transportation Engineering*, vol. 135, no. 1, pp. 9–17, 2009.
- [11] K. P. Kang and G. L. Chang, "Lane-based dynamic merge control strategy based on optimal thresholds for highway work zone operations," *Journal of Transportation Engineering*, vol. 135, no. 6, pp. 359–370, 2009.
- [12] S. Wolfram, *Theory and Applications of Cellular Automata*, World Scientific, Singapore, 1986.
- [13] S. Wolfram, *A New Kind of Science*, Wolfram Media, Champaign, Ill, USA, 2002.
- [14] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *Journal de Physique*, vol. 2, no. 12, pp. 2221–2229, 1992.
- [15] N. Duffield, C. Lund, and M. Thorup, "Estimating flow distributions from sampled flow statistics," *IEEE/ACM Transactions on Networking*, vol. 13, no. 5, pp. 933–946, 2005.
- [16] W. H. Wang, W. Zhang, H. W. Guo, H. Bubb, and K. Ikeuchi, "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [17] W. Wang, F. Hou, H. Tan, and H. Bubb, "A framework for function allocations in intelligent driver interface design for comfort and safety," *International Journal of Computational Intelligence Systems*, vol. 3, no. 5, pp. 531–541, 2010.

Research Article

Emissions and Fuel Consumption Modeling for Evaluating Environmental Effectiveness of ITS Strategies

Yuan-yuan Song,^{1,2} En-jian Yao,^{1,2} Ting Zuo,¹ and Zhi-feng Lang¹

¹ State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

² MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Yuan-yuan Song; 11120986@bjtu.edu.cn

Received 8 September 2012; Revised 14 December 2012; Accepted 15 December 2012

Academic Editor: Wuhong Wang

Copyright © 2013 Yuan-yuan Song 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.

Road transportation is a major fuel consumer and greenhouse gas emitter. Recently, the intelligent transportation systems (ITSs) technologies, which can improve traffic flow and safety, have been developed to reduce the fuel consumption and vehicle emissions. Emission and fuel consumption estimation models play a key role in the evaluation of ITS technologies. Based on the influence analysis of driving parameters on vehicle emissions, this paper establishes a set of mesoscopic vehicle emission and fuel consumption models using the real-world vehicle operation and emission data. The results demonstrate that these models are more appropriate to evaluate the environmental effectiveness of ITS strategies with enough estimation accuracy.

1. Introduction

Transportation sector has been facing an increasing environmental pressure due to the rapid motorization in China over the recent years. Low-carbon transportation solutions such as substitution of fossil oil by alternative fuels, enhancing vehicle technology, and developing public transport systems have been applied widely. An alternative and promising solution is the implementation of ITS that can smooth the traffic flow and reduce congestion. In addition to the common effects, increasing attention has been paid to the indirect effect of ITS technologies on reducing the fuel consumption and CO₂ emissions. Many researches have demonstrated the environmental improvement of ITS technologies [1, 2] especially in car navigation systems including Ecodriving [3, 4] and Ecorouting [5, 6]. Therefore, it is of significance to establish emission estimation models to evaluate the environmental effectiveness of ITS strategies with enough estimation accuracy.

Vehicle emission estimation models play a critical role for regional planning and development of emission control strategies [7]. Three general approaches are usually

considered in modeling vehicle emissions and fuel consumption [8]. Macroscopic models use average aggregate network parameters to estimate emission inventories for large regional areas according to the relationships between speed, flow, and density of a stream. Examples of macroscopic models applicable to vehicles include the U.S. federal's MOBILE6 [9] and California's EMFAC [10]. But these models are not well suited for evaluating traffic operational improvements that can be achieved through the ITS strategies [7]. Microscopic models such as VT-Micro model [11] and CMEM [12] model estimate instantaneous vehicle emission rates using either vehicle engine or vehicle speed/acceleration data. However, it is difficult to obtain substantial amounts of microscopic parameters for evaluating the environmental effectiveness of ITS strategies. Most existing road surveillance systems such as loop, remote traffic microwave sensor (RTMS), and floating car data (FCD) system, can readily acquire average link travel speed or point speed data. Given that macroscopic models ignore the transient variation of vehicle emissions associated with different traffic conditions, while microscopic emission estimation tools need numerous input data, it is more appropriate to utilize link-based mesoscopic emission

models for evaluating the environmental impacts of local ITS strategies.

There are many new approaches including MOVES to estimating the emissions with an attempt to replace existing emission models such as MOBILE [13]. MOVES incorporates the concept of vehicle specific power (VSP) and characterizes vehicle activities according to VSP and speed. VSP is defined as the instantaneous power per unit mass of the vehicle, and many studies have found methods based on VSP more accurate to estimate vehicle fuel consumption and emissions. The VSP-based methods were further developed and better accepted by other researchers in the area of emission and fuel consumption modeling [14] after its first application [15]. The ability of VSP-based emission models to reflect the transient emission rates under different operating modes as well as the better estimation of vehicle emissions than speed-based emission models makes it widely utilized. In order to evaluate the effects of traffic management on fuel efficiency, Song et al. [16] developed a practical model that aggregated the normalized fuel rate under different VSP bins. Wang and colleagues [17] established a model based on VSP and speed, which provided insight that how different levels of cruise speed and acceleration affect vehicle fuel consumption. Furthermore, the accuracy was improved with prediction errors within 20% for trip emissions and link-speed-based emission factors through follow-up studies [18]. Scora et al. [19] combined real-time traffic data along with the comprehensive modal emission models (CMEM and EPA's MOVES) to estimate the environmental measures in real time. The estimation methodology provided far more dynamic and accurate environmental information compared to static emission models.

In order to improve both the applicability and accuracy of the evaluation effect, link-based emission and fuel consumption models are necessary along with consideration of transient vehicle behavior in these models. Also, mesoscopic driving parameters collected by probe vehicle systems are well applied in some ITS strategies such as advanced traffic monitoring and management systems [20] and navigation systems [5, 6, 21]. Many existing studies have focused on vehicle emission estimation and fuel consumption models, some of which have been applied for evaluation of environmental effects [8]. However, the existing evaluation is generally realized through direct application of regulation and test procedures for fuel economy of automobiles, and the indicator is fuel consumed per unit distance such as per 100 km other than realistic operational modes. The input data of these models are generally average speed of a whole trip or a period time such as one hour, and the time ranges are too wide to take full advantage of the probe vehicle technology or other traffic information systems. Therefore, the primary objective of this paper is to develop practical emission and fuel consumption estimation models for evaluation of the environmental effectiveness of ITS strategies for improving regional traffic flow.

In Section 1 of the paper, background information on emissions and fuel consumption modeling for evaluating environmental effectiveness of ITS strategies is provided, along with a review of exiting studies on emission and

fuel consumption models. Section 2 outlines the overall methodology of the research, and the results of emissions and fuel consumption models are illustrated. These models for evaluating environmental effectiveness of ITS strategies (i.e., ecological route navigation) are applied in Section 3. Conclusions and future work are then provided in Section 4.

2. Model Estimation

Light-duty vehicles and heavy-duty vehicles account for most of the vehicle fleet in the city. ITS strategies are more widely applied in these types of vehicles. Therefore, based on the emission data collected by Portable Emission Measurement System (PEMS), mesoscopic models for light-duty gasoline vehicles and heavy-duty diesel vehicles are established in this paper. In order to evaluate the environmental effectiveness of ITS strategies that can smooth the regional traffic flow, the established vehicle emission and fuel consumption models utilize average link speed as explanatory variables. Also, the VSP distribution for each travel speed level is considered as a bridge between the instantaneous driving parameters such as vehicle speed, acceleration and average link speed, which guarantees the estimation accuracy and applicability of the models. The methodology of estimation models is listed in the following sections.

2.1. Source of Data. We use the data collected by vehicle with PEMS in Beijing urban areas in this study. The vehicles were operated on regular routes in the urban area under different driving conditions (i.e., different road grades and different traffic status). The driving conditions are different with speed from 0 to 100 km/h associated with acceleration ranging from -5 m/s^2 to 5 m/s^2 , while the corresponding VSP calculated using second-by-second data is between -30 kw/t and 25 kw/t .

2.2. Model Methodology. VSP integrates the vehicle speed, vehicle acceleration, road grade, aerodynamic drag, and tire rolling resistance, and it is generally defined as the instantaneous power per unit mass of the vehicle [22]. In this paper, simplified formulas are used to calculate the values of VSP for different types of vehicles. The following simplified function [15] in (1) is used for the VSP calculations of light-duty gasoline vehicles while heavy-duty diesel vehicle's VSP can be calculated using (2) according to the existing research results [23]:

$$\text{VSP} = v \times (1.1 \times a + 0.132) + 0.000302 \times v^3, \quad (1)$$

$$\text{VSP} = v \times (a + 0.09199) + 0.000169 \times v^3, \quad (2)$$

where v and a are the vehicle speed and acceleration in m/s and m/s^2 , respectively.

For the characteristics of vehicle emission change greatly under different travel conditions, VSP is separated as a bin with an equal interval of 1 kw/t , as described in

$$\forall : \text{VSP} \in [n, n + 1), \quad \text{VSP bin} = n, \quad n \text{ is integer.} \quad (3)$$

Each VSP bin is associated with average emission rates for different types of emissions, respectively. The second-by-second data collected by PEMS is divided into traveling fragment by different time interval and the optimal time granularity is discussed in the following section. Each fragment is characterized with its average speed that is as the basis of division of fragments. After the interval division of average speed, the VSP-Bin distribution attribute of each average speed range is calculated. The average emission rate under each average speed range is estimated as

$$\overline{ER}_i = \sum_j ER_j \times \frac{t_{i,j}}{T_i}, \quad (4)$$

where \overline{ER}_i is the average emission rate under average speed range i , g/s; j is the index of VSP bin; ER_j is the emission rate for VSP bin j , g/s; $t_{i,j}$ is the time spent in VSP bin j on speed range i , s; and T_i is the total travel time under speed range i , s.

The instantaneous fuel consumption is calculated from emissions of CO₂, HC, and CO using carbon balance method listed in national standards of China [24]. Vehicle fuel consumption rates for gasoline vehicles and diesel vehicles can be estimated using (5) and (6), respectively, as follows:

$$FR_S = 1.154 \times \left(ER_{HC} \times \frac{12}{13} + ER_{CO} \times \frac{12}{28} + ER_{CO_2} \times \frac{12}{44} \right), \quad (5)$$

$$FR_C = 1.155 \times \left(ER_{HC} \times \frac{12}{13} + ER_{CO} \times \frac{12}{28} + ER_{CO_2} \times \frac{12}{44} \right), \quad (6)$$

where FR_S and FR_C are the fuel consumption rates of gasoline vehicles and diesel vehicles, respectively, g/s; ER_{HC} , ER_{CO} , and ER_{CO_2} are the HC, CO, and CO₂ emission rates, respectively, g/s.

Then the emission/fuel consumption factor under each average speed range is estimated as:

$$V_i = \frac{\sum_{k=1}^n D_k}{\sum_{k=1}^n T_k}, \quad (7)$$

$$EF_i (FF_i) = ER_i (FR_i) \times \frac{\sum_{k=1}^n T_k}{\sum_{k=1}^n D_k},$$

where T_k is the vehicle trip time spent in the travel fragment k of average speed range i , s; D_k is the vehicle trip distance in travel fragment k of average speed range i , km; V_i is the vehicle travel speed for average speed range i , km/h; $ER_i (FR_i)$ is the emission (fuel consumption) rate for average speed range i , g/s; and $EF_i (FF_i)$ is the emission (fuel consumption) factor for V_i .

Based on the previous research [25] and the relationship of emission (fuel consumption) factors versus average speed, (8) is used as the fitted formula between vehicle emission (fuel consumption) factor s and average speed:

$$EF (FF) = \frac{a}{v} + b + cv + dv^2, \quad (8)$$

where $EF (FF)$ is the emission (fuel consumption) factor, g/km; v is average speed, km/h; a , b , c , and d are coefficients.

2.3. Estimation Results. Figure 1 shows the similar characteristic of various vehicle emission and fuel consumption rates for light-duty and heavy-duty as VSP changes and a VSP of 0 kW/t is the inflection point. When the VSP value is positive, emission and fuel consumption rates indicate a rapid increase with the increase of VSP. Emission and fuel consumption rates tend to be very low and almost invariable for negative VSP.

The emission and fuel consumption rates of two types of vehicles are illustrated in Figure 2. As illustrated in Figure 2, the emission and fuel consumption rates typically increase with the increment of average speed. It should be noted that there is an apparent increase for emission and fuel consumption rates during the lower average speed range, while the rates rise relatively slowly when the speed increase to a specific value.

The emission and fuel consumption models have been established based on the proposed approach. As the traveling fragment by different time interval has effects on the model precision and estimation errors. It is valuable to explore the least estimation error and discuss the optimal time granularity. Successive 600-second-long measurement trips for different time granularities were used for validation. The errors between the modeled and measured trip emissions and fuel consumption are shown in Table 1.

Table 1 illustrated that the modeled and measured emission and fuel consumption rates for light-duty vehicle under different time granularities are in good agreement, and all the differences between them are within 10%. Moreover, the time granularity has a less significant effect on the estimation errors. The optimal time granularity can be regarded as 60 second.

The vehicle emission and fuel consumption factor curves and estimation models for both vehicle types are shown in Figure 3 and Table 2 when the time granularity is 60 second.

Figure 3 illustrates that the changing tendencies for all vehicle emission curves are consistent with the exiting research results [11, 26, 27]: vehicle emission factors decrease as speed increases to a specified value, and then start to increase. Noteworthy is the fact that the values of the emission and fuel consumption factors drop dramatically at a lower average speed. For example, for light-duty vehicle, the value of CO₂ emission per kilometer declines greatly with the increase of speed, once the speed reaches to about 65 km/h, the value of CO₂ emission per kilometer will increase slowly. The optimal average speed is approximately 65 km/h with the minimum CO₂ emission rate of 210 g/km.

3. Application

Many ITS strategies have indirect influence on reducing emissions and fuel consumption through standardizing driving behavior, smoothing road traffic flow, and improving commuting efficiency. For example, the technology of traffic control in merging areas can provide a proper merging speed and opportunity for vehicles, which leads to improvement of

TABLE 1: Comparison of emission and fuel consumption rates from modeled and measured values.

	Error (%) for different time granularities					
	40 s	60 s	80 s	100 s	120 s	300 s
HC	1.99	1.27	2.93	3.42	3.39	3.93
NO _x	3.88	0.70	1.79	9.11	6.18	9.97
CO	4.21	3.22	2.72	3.56	5.28	6.27
CO ₂	1.33	2.05	2.71	3.04	4.35	5.21
Fuel	1.43	2.08	2.71	3.06	4.37	6.25

TABLE 2: Vehicle emission and fuel consumption factor models.

Vehicle type	Estimation models
Light duty vehicles	$EF_{\text{HC}} = 1.08 \times 10^1 \times V^{-1} - 7.11 \times 10^{-3} + 3.76 \times 10^{-4} \times V + 3.63 \times 10^{-5} \times V^2, R^2 = 0.91$
	$EF_{\text{NO}_x} = 2.00 \times V^{-1} - 4.49 \times 10^{-2} - 3.36 \times 10^{-4} \times V + 3.49 \times 10^{-5} \times V^2, R^2 = 0.87$
	$EF_{\text{CO}} = 8.08 \times 10^1 \times V^{-1} + 1.16 + 5.03 \times 10^{-3} \times V + 5.35 \times 10^{-4} \times V^2, R^2 = 0.94$
	$EF_{\text{CO}_2} = 4.78 \times 10^3 \times V^{-1} + 1.11 \times 10^2 - 1.24 \times V + 2.37 \times 10^{-2} \times V^2, R^2 = 0.95$
	$FF = 1.56 \times 10^2 \times V^{-1} + 3.54 - 3.88 \times 10^{-2} \times V + 7.76 \times 10^{-4} \times V^2, R^2 = 0.95$
Heavy duty vehicles	$EF_{\text{HC}} = 1.55 \times 10^1 \times V^{-1} + 3.92 \times 10^{-1} - 7.20 \times 10^{-3} \times V + 5.31 \times 10^{-5} \times V^2, R^2 = 0.98$
	$EF_{\text{NO}_x} = 8.91 \times 10^1 \times V^{-1} + 9.35 - 1.36 \times 10^{-1} \times V + 8.91 \times 10^{-4} \times V^2, R^2 = 0.98$
	$EF_{\text{CO}} = 4.14 \times 10^1 \times V^{-1} + 1.99 - 1.10 \times 10^{-2} \times V + 2.99 \times 10^{-5} \times V^2, R^2 = 0.99$
	$EF_{\text{CO}_2} = 3.67 \times 10^3 \times V^{-1} + 5.34 \times 10^2 - 7.90 \times V + 5.43 \times 10^{-2} \times V^2, R^2 = 0.99$
	$FF = 1.19 \times 10^2 \times V^{-1} + 1.69 \times 10^1 - 2.50 \times 10^{-1} \times V + 1.72 \times 10^{-3} \times V^2, R^2 = 0.99$

Note: EF is the emission factor, g/km; FF is the fuel consumption factor, g/km; V is average speed, km/h.

travel speed for upstream traffic flow. The proposed vehicle emission and fuel consumption factor models are described as functions of average link speed that can be collected by most existing road traffic information systems such as loop-coil detectors. Further, the effectiveness of emission and fuel consumption reduction utilizing the technology of traffic control can be evaluated through these emission and fuel consumption models. Moreover, the models developed in this study are incorporated into an ecological route navigation system to further demonstrate their applicability to ITS strategies.

The studying area is located in central area of Beijing. Based on the route planning algorithm, the ecological route navigation system consists of a dynamic traffic information database, emissions/fuel estimation models and user interfaces, which can provide vehicle the least emissions or fuel. The dynamic traffic information database is collected from a probe vehicle system, which provides travel time for each link every five intervals. It means only average travel speed of each link can be obtained and used for the ecological route navigation.

Figure 4 illustrates the route navigation results of the ecological route compared with the time priority route for the specified OD pair. It is obvious that ecological route is different from the time priority route. Based on the proposed emissions and fuel consumption models, the comparison results between the two routes are summarized in Table 3. Values in the table are normalized to the time priority route's

results. The fuel consumption of the ecological route is about 13.5% lower than that of the time priority route, although the travel time of the ecological route is just about 1.2% longer. Noteworthy is the fact that the value of the fuel consumption reduction is extremely similar in the appearance to the CO₂ reduction. This result verifies that estimation models can be utilized to calculate the emissions and fuel consumption during the whole trip and evaluate the environmental effect on emissions and fuel consumption reduction.

4. Conclusions

The paper presents a methodology for establishing mesoscopic emission and fuel consumption models for assessing the environmental impacts of ITS strategies. The proposed models are developed through considering the influence of the vehicle's operating mode on vehicle emissions, which not only guarantees the accuracy of emissions and fuel consumption models, but also makes it possible to estimate the emission and fuel consumption based on most current traffic information systems. With more and more PEMS data for specific vehicles are obtained in the future, the models for other types of vehicles can be estimated and the emissions inventory is expected to be accomplished. Furthermore, it is verified that these models are well applied to evaluate the effect of ITS technologies on reducing vehicle emissions and fuel consumption.

TABLE 3: Results of evaluation of ecological route strategy.

	Ecological route	Time-priority route	% Differences
Travel distance (km)	13.3	16.2	-17.9%
Travel time (min)	17.6	17.4	1.2%
CO ₂ emission (kg)	3.2	3.6	-13.5%
Fuel consumption (kg)	1.0	1.2	-13.5%

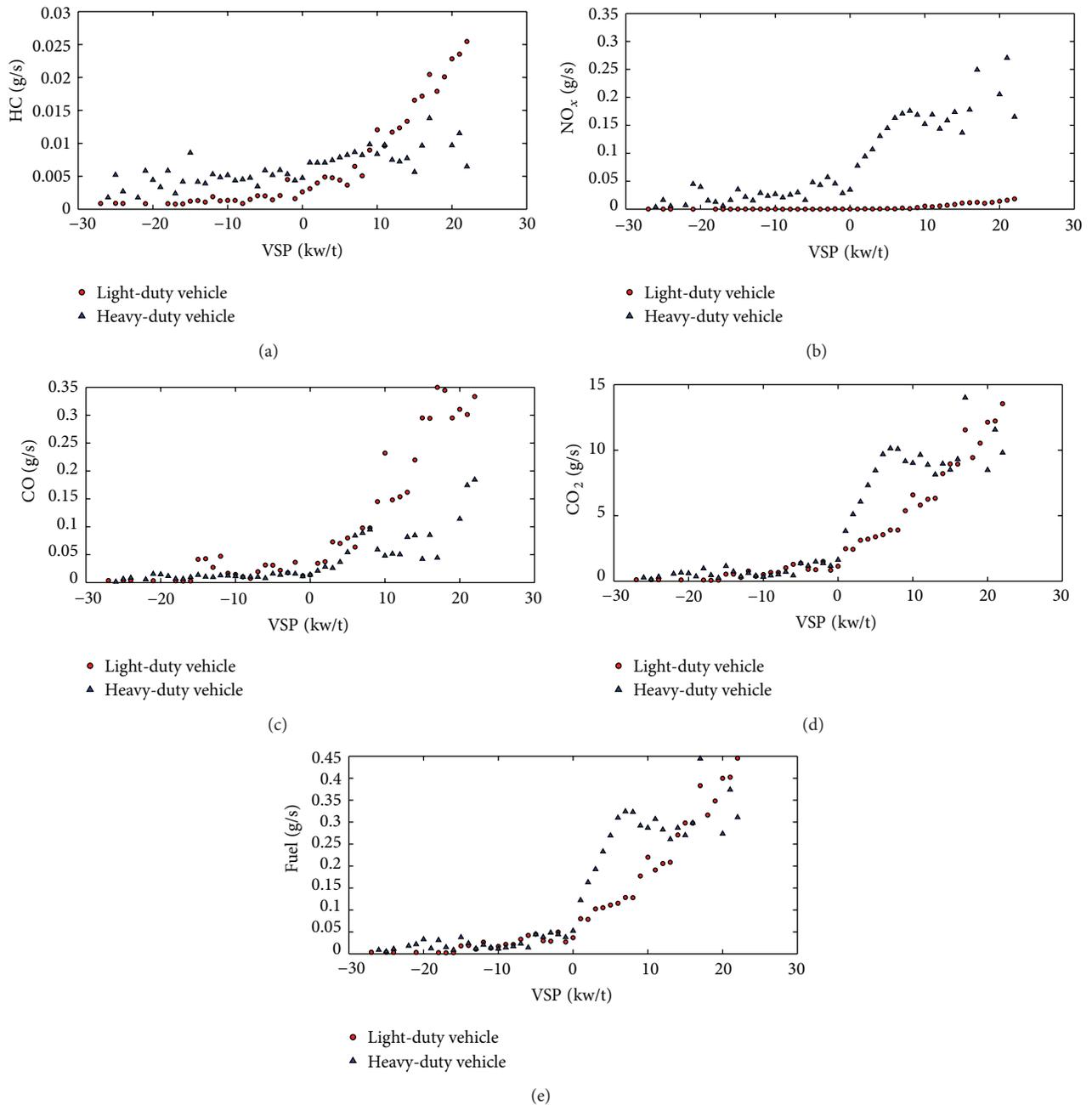


FIGURE 1: Emission and fuel consumption rates under different VSP-bins (Light- and heavy-duty vehicle).

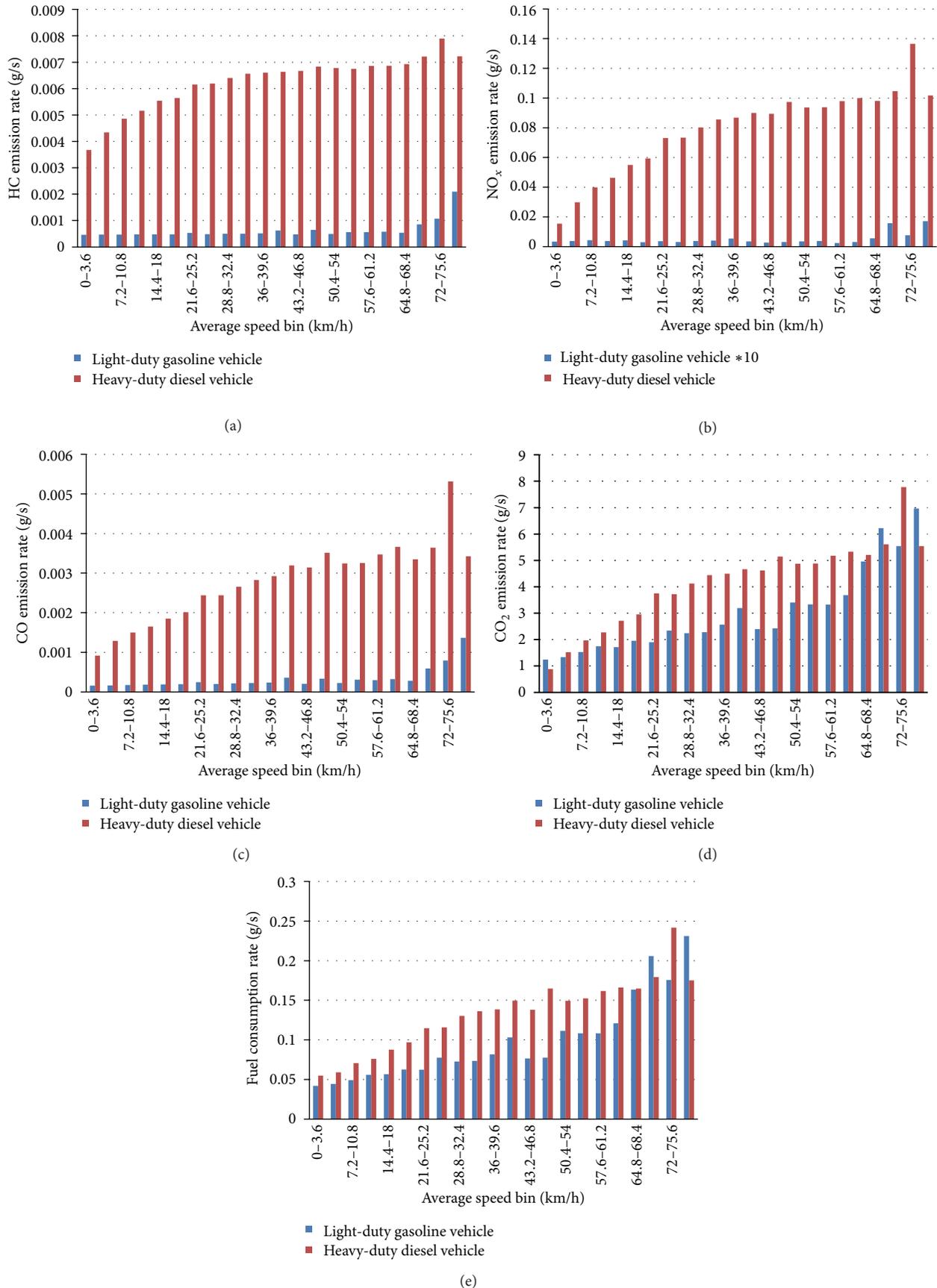
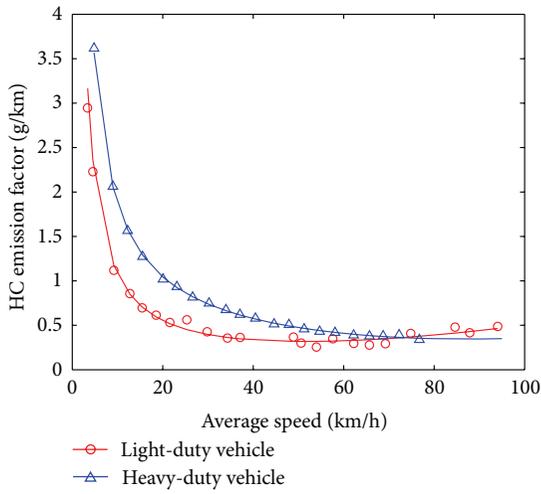
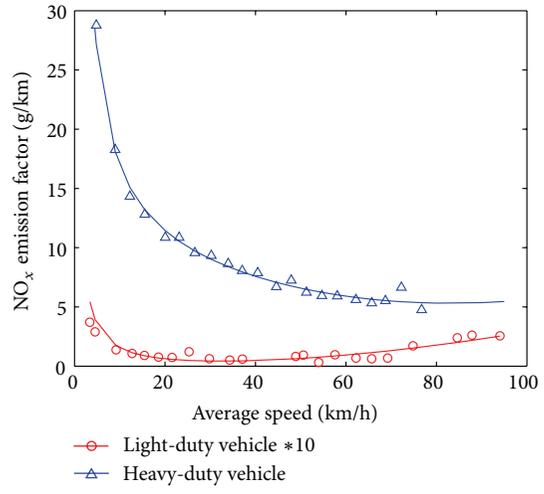


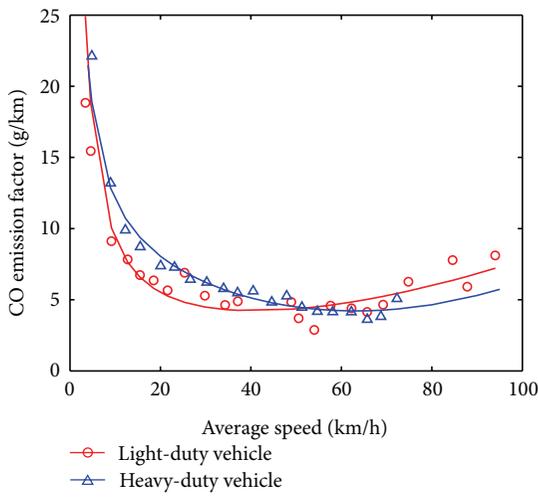
FIGURE 2: Emission and fuel consumption prediction values under different average speed-bins (light- and heavy-duty vehicle).



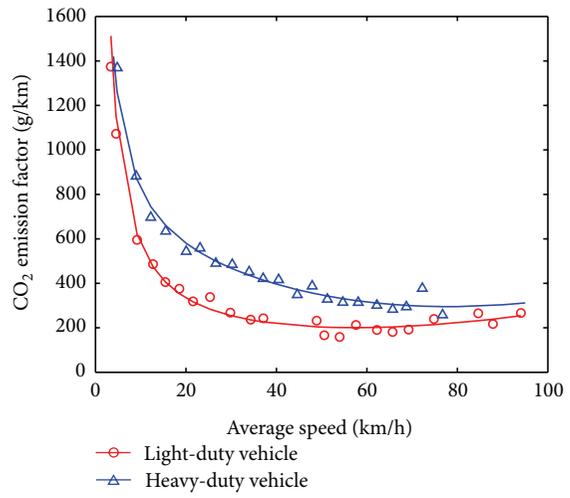
(a)



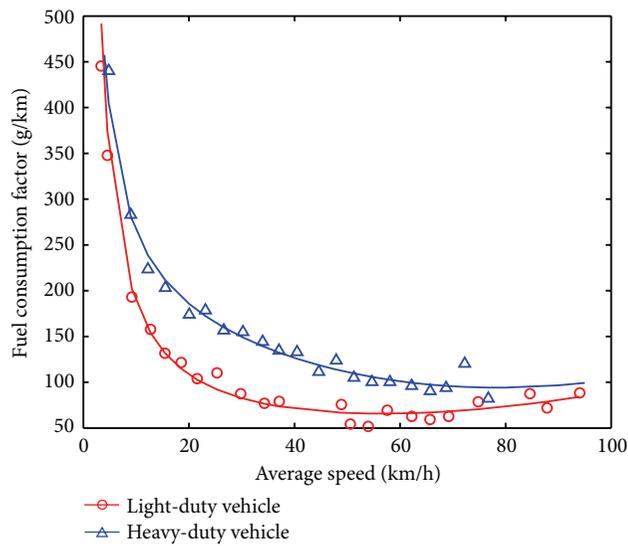
(b)



(c)



(d)



(e)

FIGURE 3: Vehicle emission and fuel consumption curves (light- and heavy-duty vehicle).



FIGURE 4: Time priority route and ecological route.

Acknowledgment

This work is supported by National 973 Program of China (no. 2012CB725403).

References

- [1] eCoMove Consortium, *ECoMove—Description of Work*, eCoMove Consortium, Brussels, Belgium, 2010.
- [2] J. D. Vreeswijk, M. K. M. Mahmood, and B. van Arem, “Energy efficient traffic management and control—the eCoMove approach and expected benefits,” in *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC ’10)*, pp. 955–961, Madeira, Portugal, September 2010.
- [3] M. van der Voort, *Design and evaluation of a new fuel-efficiency support tool [Ph.D. thesis]*, University of Twente, 2001.
- [4] E. Ericsson, H. Larsson, and K. Brundell-Freij, “Optimizing route choice for lowest fuel consumption—potential effects of a new driver support tool,” *Transportation Research C*, vol. 14, no. 6, pp. 369–383, 2006.
- [5] 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.
- [6] 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.
- [7] M. Barth, C. Malcolm, T. Younglove, and N. Hill, “Recent validation efforts for a comprehensive modal emissions model,” *Transportation Research Record*, no. 1750, pp. 13–23, 2001.
- [8] H. Dia, S. Panwai, N. Boongrapue, T. Ton, and N. Smith, “Comparative evaluation of power-based environmental emissions models,” in *Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC ’06)*, pp. 1251–1256, Toronto, Canada, September 2006.
- [9] U.S. Environmental Protection Agency, “User’s guide to MOBILE6. 1 and MOBILE6. 2: mobile source emission factor model,” Report No. EPA420-R-03-010, U.S. Environmental Protection Agency, Ann Arbor, Mich, USA, 2003.
- [10] California Air Resources Board, *EMFAC, 2007 Version 2.30 User’s Guide: Calculating Emission Inventories for Vehicles in California*, California Air Resources Board, Sacramento, Calif, USA, 2007.
- [11] K. Ahn, A. A. Trani, H. Rakha, and M. van Aerde, “Microscopic fuel consumption and emission models,” in *Proceedings of the 78th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 1999.
- [12] M. Barth, F. An, T. Younglove et al., *Comprehensive Modal Emission Model (CMEM), Version 2.0 User’s Guide*, 2000.
- [13] J. Koupal, H. Michaels, M. Cumberworth, C. Bailey, and D. Brzezinski, “EPA’s plan for MOVES: a comprehensive mobile source emissions model,” USEPA Documentation, 2002.
- [14] E. K. Nam and R. Giannelli, “Fuel consumption modeling of conventional and advanced technology vehicles in the physical emission rate estimator (PERE),” Draft Report No. EPA420-P-05-001, U.S. Environmental Protection Agency, Washington, DC, USA, 2005.
- [15] J. Jiménez-Palacios, *Understanding and quantifying motor vehicle emissions with vehicle specific power and TILDS remote sensing [Ph.D. thesis]*, Massachusetts Institute of Technology, 1999.
- [16] G. Song, L. Yu, and Z. Wang, “Aggregate fuel consumption model of light-duty vehicles for evaluating effectiveness of traffic management strategies on fuels,” *Journal of Transportation Engineering*, vol. 135, no. 9, pp. 611–618, 2009.
- [17] 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.
- [18] H. Wang and L. Fu, “Developing a high-resolution vehicular emission inventory by integrating an emission model and a traffic model: part I—modeling fuel consumption and emissions based on speed and vehicle-specific power,” *Journal of the Air and Waste Management Association*, vol. 60, no. 12, pp. 1463–1470, 2010.
- [19] G. Scora, B. Morris, and C. Tran, “Real-time roadway emissions estimation using visual traffic measurements,” in *Proceedings of the IEEE Forum on Integrated and Sustainable Transportation Systems*, 2011.
- [20] *PeMS Performance Measurement System 10.3*. Caltrans, The University of California, PATH, Berkeley, Calif, USA, 2010.
- [21] W. Wang, *Vehicle’s Man-Machine Interaction Safety and Driver Assistance*, China Communications Press, Beijing, China, 2012.
- [22] H. C. Frey, N. M. Roupail, H. Zhai, T. L. Farias, and G. A. Gonçalves, “Comparing real-world fuel consumption for diesel- and hydrogen-fueled transit buses and implication for emissions,” *Transportation Research D*, vol. 12, no. 4, pp. 281–291, 2007.
- [23] P. Andrei, *Real world heavy-duty vehicle emission modeling [M.S. thesis]*, University of West Virginia, 2001.
- [24] CN-GB, “Measurement methods of fuel consumption for light-duty vehicles,” 2008.
- [25] Y. Namikawa, Y. Takai, and N. Ohshiro, *Calculation Vase of Motor Vehicle Emission Factors*, National Institute for Land and Infrastructure Management, Japan, 2003.
- [26] M. André and M. Rapone, “Analysis and modelling of the pollutant emissions from European cars regarding the driving

characteristics and test cycles,” *Atmospheric Environment*, vol. 43, no. 5, pp. 986–995, 2009.

- [27] S. K. Zegeye, B. de Schutter, H. Hellendoorn, and E. Breunese, “Model-based traffic control for balanced reduction of fuel consumption, emissions, and travel time,” in *Proceedings of the 12th IFAC Symposium on Transport Systems (CT '09)*, pp. 149–154, Redondo Beach, Calif, USA, September 2009.

Research Article

A Real-Time Lane Detection Algorithm Based on Intelligent CCD Parameters Regulation

**Ping-shu Ge,¹ Lie Guo,² Guo-kai Xu,¹
Rong-hui Zhang,³ and Tao Zhang¹**

¹ College of Electromechanical and Information Engineering, Dalian Nationalities University, Dalian, Liaoning 116600, China

² School of Automotive Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

³ Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China

Correspondence should be addressed to Lie Guo, guolie@163.com

Received 27 August 2012; Revised 23 November 2012; Accepted 27 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Ping-shu Ge 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.

Lane departure warning system (LDWS) has been regarded as an efficient method to lessen the damages of road traffic accident resulting from driver fatigue or inattention. Lane detection is one of the key techniques for LDWS. To overcome the contradiction between complexity of algorithm and the real-time requirement for vehicle onboard system, this paper introduces a new lane detection method based on intelligent CCD parameters regulation. In order to improve the real-time capability of the system, a CCD parameters regulating method is proposed which enhances the contrast between lane line and road surfaces and reduces image noise, so it lays a good foundation for the following lane detection. Hough transform algorithm is improved by selection and classification of seed points. Finally the lane line is extracted through some restrictions. Experimental results verify the effectiveness of the proposed method, which improves not only real-time capability but also the accuracy of the system.

1. Introduction

With the rapid development of expressway and the growth of motor vehicle quantities, the traffic accidents, especially mass traffic accidents, are increasing and become one of the most serious problems around the world [1–4]. For low-income and middle-income countries, this situation is much worse. A report from World Health Organization (WHO) points out that over 90% of the world's fatalities on the roads occur in low-income and middle-income countries, which have only 48% of the world's registered vehicles. It has been estimated that, unless immediate action is taken, road traffic deaths will rise to the fifth leading cause of

death by the year 2030, resulting in an estimated 2.4 million fatalities per year [5]. The global status report on road safety clearly shows that significantly more action is needed to make the world's roads safer. At present, the increase trend of traffic accidents and casualties in China has been slowed down to some extent and shows a declining tendency, but the situation is far away from being optimistic [6].

To lessen the damages resulting from the road traffic accidents and save more lives, the reasons and factors should be analyzed and summarized. There are a number of factors which cause the traffic accidents, ranging from driver behavior to mechanical failure, environmental conditions, and roads design [7–9]. Many works have been done to investigate ways of developing a vehicle driving assistance system to improve its safety. Among those factors resulting in serious accidents, driver fatigue or inattention contribute a considerable portion [10, 11]. Driving is a complex process which involves eye-hand-foot coordination. In many cases, the driver falls asleep making the vehicle to leave its designated lane and possibly causing an accident. NHTSA estimated that running off the road caused about 28% of fatal vehicle accidents in the US in 2005. Moreover, drowsy, sleeping, or fatigued and inattentive drivers caused about 2.6% and 5.8% of the fatal crashes, respectively [12, 13]. Further research is needed to improve driver's behavior and driving activities with the help of other advanced techniques.

In order to prevent this type of accident, researchers have proposed a variety of solutions and technologies to predict and detect unintended lane departure events to warn drivers about such events. Lane departure warning system (LDWS) has been proposed for this purpose, to warn the driver as soon as a vehicle begins to inadvertently drift out of its driving lane or to automatically take steps to ensure that the vehicle stays in correct lane [14]. The currently available LDWSs on the markets are forward-looking vision-based systems and are mainly applied to prestige car, such as BMW, Mercedes-Benz, and so on. In order to reduce the traffic accidents, the LDWS needs to adapt all types of vehicle rather than some certain kinds of vehicles. Although many researchers have proposed several advanced algorithms to improve the system performance [15, 16], there still exist the following bottleneck problems:

- (1) Lane detection should adapt to various conditions. Lane detection is a main task of LDWS, and the accuracy of detection is foundation and prerequisite to realize its warning function. Several algorithms have been proposed and their differences are mainly consisting in image preprocessing, lane model, selected model fitting method, and tracking strategy [17, 18]. However, it is difficult to be performed with a high detection rate in complex situations involving shadows, varying illumination conditions, bad conditions of road paintings and kinds of lane marks such as solid lines, segment lines, double yellow lines, pavement or physical barriers. Therefore, how to make the detection algorithm accommodate the complex environment is an important issue.
- (2) The LDWS has to meet the requirements such as robust, low cost, compact, low dissipation, and real time. Some of these requirements need to be accomplished by algorithm, while the other can be achieved from the hardware device. In [19], an adaptive DSP-based LDWS was introduced with its operating frequency of 600 MHz and the lane marking detection speed of 35 f/s. The system function is compact but the price is slightly high. In addition, researchers use FPGA devices or other low-cost and low-power consumption architectures to develop the systems [20, 21]. However, the contradiction between complexity of algorithm and the memory of hardware is very outstanding.

To overcome the above difficulty, some strategies should be developed to make the system real time and robust, which can adapt to real road environments. This paper aims to develop a real-time lane detection algorithm based on intelligent regulated CCD parameters, which can effectively solve the conflicts between real time and robustness of the LDWS. The contents of this paper are organized as follows. Section 2 introduces an intelligent regulated CCD parameters algorithm. After analyzing the advantages and disadvantages of traditional Hough transform, an improved Hough transform algorithm is proposed in Section 3. Section 4 verifies the effectiveness of the proposed algorithm through road experiments, and some conclusions are given in Section 5.

2. CCD Parameters Regulation Based on Feature Regions

The LDWS system needs a great demand on the CCD image grabbing speed. To achieve this, CCD employs a kind of parameter control technique, that is to say, the gain, bright, and shutter of CCD can be programmed. Therefore, the CCD parameters can be regulated by programmed in real time, which is helpful to image segmentation, reducing noise as well as improving the reliability and instantaneity of lane detection algorithms. Moreover, it can heighten the contrast between road surfaces and lane marks.

The CCD model used in our LDWS is BaslerA602f and its parameters can be regulated through IEEE1394 video capture board connected to an imbedded computer [22]. The purpose of CCD parameters regulation can be realized by defining some feature regions and calculating the eigenvalues of those regions.

2.1. Feature Regions Division and Eigenvalues Computation

As showed in Figure 1, the image to be processed is divided into six feature regions from region A to F. The size of image is 320×240 pixels.

Region A is the upper center 1/4 region of the image. The gray mean value μ_a and variance σ_a of this region indicate the brightness of sky. While region B is the bottom center 1/4 region of the image, whose gray mean value μ_b and variance σ_b indicate the darkness of the road surface. They can be calculated using

$$\begin{aligned}
 \mu_a &= \frac{32}{3 \times w \times h} \times \sum f(x, y) & -\frac{3w}{16} \leq x \leq \frac{3w}{16}, \frac{3}{4}h \leq y \leq h, \\
 \sigma_a &= \sqrt{\frac{32 \times \sum [f(x, y) - \mu_a]^2}{3 \times w \times h}} \\
 \mu_b &= \frac{32}{3 \times w \times h} \times \sum f(x, y) & -\frac{3w}{16} \leq x \leq \frac{3w}{16}, 0 \leq y \leq \frac{h}{4}, \\
 \sigma_b &= \sqrt{\frac{32 \times \sum [f(x, y) - \mu_b]^2}{3 \times w \times h}}
 \end{aligned} \tag{2.1}$$

where w and h are the width and height of image, respectively; $f(x, y)$ is the gray value of pixel (x, y) .

$(0, 120)$, $(80, 120)$ respectively. The function for PQ can be expressed by $y = x + 120$. Then the eigenvalues of regions C and D are calculated as

$$u_c = \frac{\sum_{y=0}^{h/6} \sum_{x=-w/2}^{y-120} f(x, y)}{s_1} + \frac{\sum_{y=h/6}^{h/4} \sum_{x=y-200}^{y-120} f(x, y)}{s_2}, \quad (2.2)$$

$$\sigma_c = \sqrt{\frac{\sum_{y=0}^{h/6} \sum_{x=-w/2}^{y-120} [f(x, y) - u_c]^2}{s_1}} + \sqrt{\frac{\sum_{y=h/6}^{h/4} \sum_{x=y-200}^{y-120} [f(x, y) - u_c]^2}{s_2}},$$

$$u_d = \frac{\sum_{y=h/4}^{h/2} \sum_{x=y-200}^{y-120} f(x, y)}{s_d}, \quad (2.3)$$

$$\sigma_d = \sqrt{\frac{\sum_{y=h/4}^{h/2} \sum_{x=y-200}^{y-120} [f(x, y) - u_d]^2}{s_d}},$$

where s_1 and s_2 are pixel numbers in regions PHM and $PIGH$, respectively; s_d is the pixel number in region D .

The angle between ST and x -axis is 135° and ST is parallel to NL . The gray mean values and variances of regions E and F can be calculated by the same means.

If the lane is during the tracking status, that is to say, we have gotten prior parameters of the lane to be detected. In this condition, the instantaneity of the system is improved by building some dynamic trapezoid feature regions. Due to the effect of projection, there are some differences in deflected distance between the bottom and the centre of the image. Therefore, we build the dynamic trapezoid feature areas taking advantage of 6 specialized feature points [24]. As shown in Figure 2, point O is taken as the previous lane vanishing point, with its coordinate is (O_x, O_y) . Points $L(L_x, L_y)$ and $R(R_x, R_y)$ are the endpoint of the left lane and the right lane, respectively. Then offset point O with 20 pixels to the left and right, respectively, we can get points $L_1(O_x-20, O_y)$ and $R_1(O_x+20, O_y)$. Offset point L with 30 pixels up and connect this point to L_1 , the crosspoint for this line with the image margin is L_2 . Offset point L with 30 pixels down and connect this point to R_1 , the crosspoint for this line with the image margin is L_3 . At the same way, we will get points R_2 and R_3 . In the tracking condition, a moving vehicle will change the lane position, which leads to the changing of these 6 feature points. As a result, the feature regions determined by these points are dynamic.

2.2. CCD Parameters Regulation

Once the feature regions are determined and their eigenvalues are calculated, the CCD parameters can be regulated according to those values. The detailed CCD parameters regulation is as follows.

- (1) Computing the eigenvalues of those feature regions according to the lane detection and tracking status.
- (2) Previous experiments in different illumination conditions indicate that it is easy to make a decision between lane marks and road surfaces when the mean values of

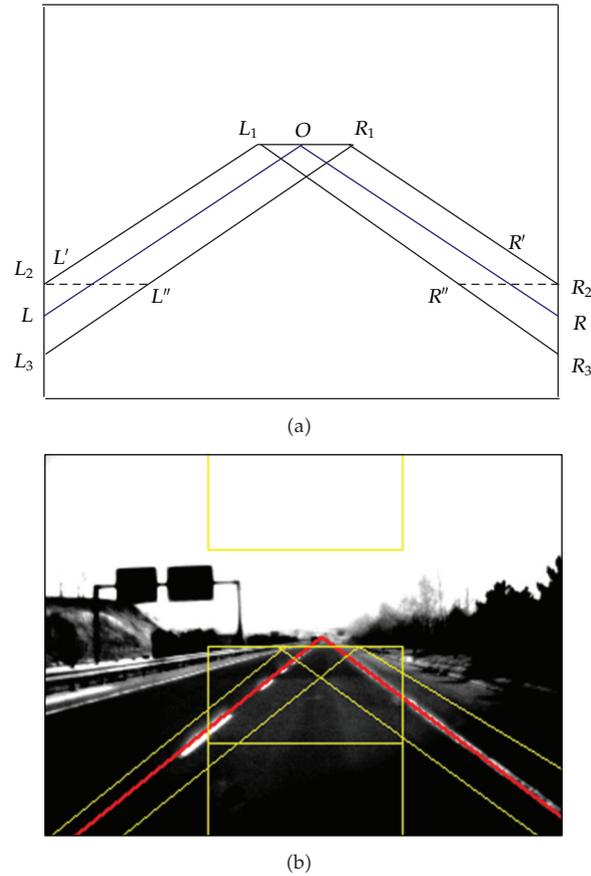


Figure 2: Schematic of feature regions division in tracking status.

regions C , D , E , and F are in range of $[60,100]$ and their variances in $[11, 27]$. So we will not change the CCD parameters and continuous capturing images when the gray level mean values gotten are in range $[60, 100]$ and their variances in $[11, 27]$. If all of the mean values are below 60 or higher than 100, we will change the CCD gain, brightness, and shutter by a certain step size till their eigenvalues are in proper range.

(3) Saving the current parameters and capture the next frame images.

Figure 3 shows the image and its segmentation results after the CCD parameters regulation in the condition of strong illumination. In these images showed below, the left one is the image based on CCD parameters regulation and the right one is the corresponding segmentation result.

Above images show that the contrast between lane marks and the road surfaces has been enhanced after the CCD parameters regulation. The segmentation of the image is pretty well and there is little background noise.

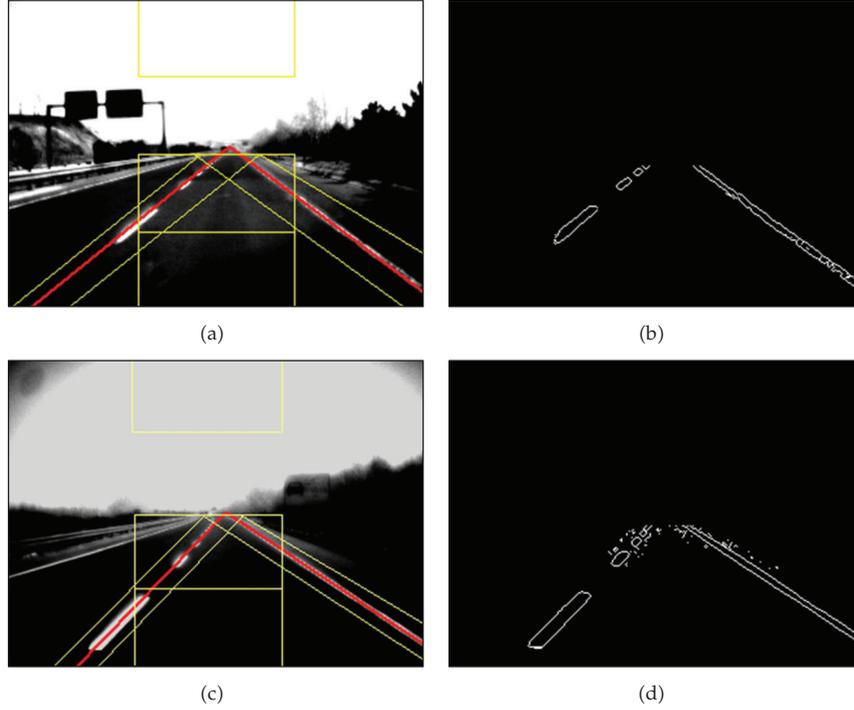


Figure 3: Image captured after the CCD parameters regulation and its segmentation result.

3. Lane Detection Using Improved Hough Transform

3.1. Analysis of Traditional Hough Transform

As the images collected on the highway are very complicated, there will be great error if we use linear fitting method to detect the lane. On the contrary, Hough transform is robust to small change of the image noise, and it is good at handling over the conditions that the object is partly occluded and covered [25]. In the image space, the straight line can be described as $y = mx + b$ and can be graphically plotted for each pair of image points (x, y) . The main idea of Hough transform is to consider the characteristics of the straight line not as image points $(x_1, y_1), (x_2, y_2), \dots$ but instead, in terms of its parameters, that is, the slope parameter m and the intercept parameter b . For computational reasons, a different pair of parameters in polar coordinates, denoted by ρ and θ , is used for the line in the Hough transform. ρ is the length of a normal from the origin to this line and θ is the orientation of ρ with respect to the x -axis. The implementation procedure using traditional Hough transform to extract a straight line is as follows.

- (1) Quantize ρ and θ , build a two-dimensional accumulate array $M(\rho, \theta)$ in the parameter map of (ρ, θ) , as shown in Figure 4. The span for parameter (ρ, θ) are separately $[\rho_{\min}, \rho_{\max}]$ and $[\theta_{\min}, \theta_{\max}]$.
- (2) Initialize the two-dimensional array $M(\rho, \theta)$, search the white edge in the image in sequence. To every white edge in the binarization image, let θ can be any value in the axis θ . Then calculate the value ρ according to the equation $\rho = x \cos \theta + y \sin \theta$. Accumulate the array M on the basis of the value of θ and ρ : $M(\rho, \theta) = M(\rho, \theta) + 1$.

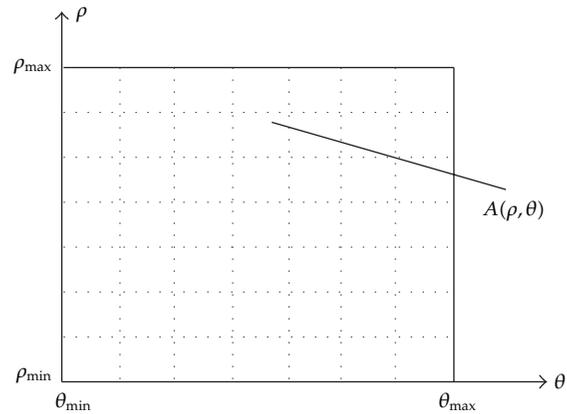


Figure 4: Accumulate array $M(\rho, \theta)$ in the parameter map of (ρ, θ) .

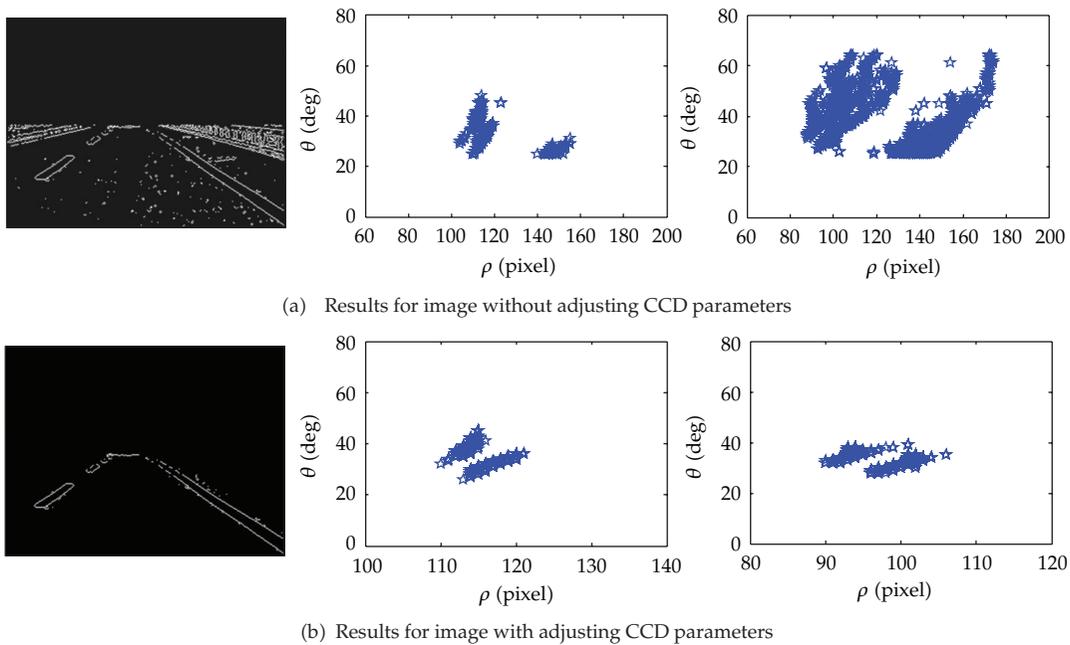


Figure 5: Peak value distributions for traditional Hough transform.

- (3) Figure out the maximum value in the transform domain and record it.
- (4) Clear the maximum value and their values of its neighborhood points.
- (5) Figure out all the maximum points successively and record them.
- (6) Seek for the recent lane line according to some certain constraints.

According to this procedure, the traditional Hough transform has some disadvantages as follows when detecting the road lane.

- (1) As every white point should be involved in the computing of space transform, the calculated amount increases, and it is time consuming. We can overcome this

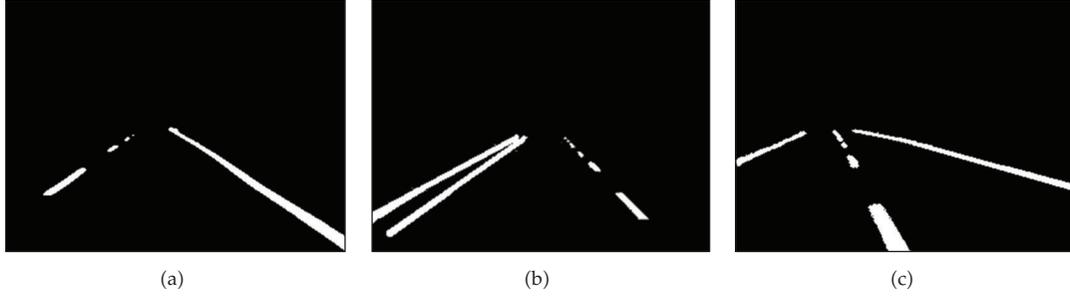


Figure 6: Road images after threshold and noise filtering. (a) Normal condition, (b) existing interference, (c) Deviation happens.

by adjusting the CCD parameters, as shown in the Figure 5. We compare the lane detection using the binarization images collected at the same place. The left images in Figure 5 show the situation without and with adjusting CCD parameters, respectively. Through comparing these two figures, we discover that after the use of CCD parameter adjusting algorithm, the points involving in the Hough transform decrease clearly, which verifies the effectiveness of CCD parameter adjusting algorithm.

- (2) Due to the quantizing of parameters ρ and θ , the peak value near one straight line is always very large after finding the peak value; we should clear the neighborhood values and look for next line. But it is not easy to perfectly define the neighborhood range, if the neighborhood is too small, the line detected next time will be overlapped on the already existed line, if the neighborhood is too large, the nearby waiting to be detected points will be clear up, and next time, you will not find any point. It is more obvious when there exist several similar lines at the same time and they are close to each other, as shown in the middle and right column of Figure 5.

3.2. Improved Hough Transform for Lane Detection

Facing the disadvantages, many researchers have proposed some improved methods. Those methods can be summarized to two kinds [26, 27]. The one is to classify the points in the image before conducting Hough transform, such as using the gradient direction to reduce the number of votes. This kind can reduce the computation time and has the interesting effect of reducing the number of useless votes, thus enhancing the visibility of the spikes corresponding to real lines in the image. The key technique is to select proper point classifying method. The other is to improve the transform voting scheme, such as Kernel-based Hough transform [28]. This approach can significantly improve the performance of the voting scheme and makes the transform more robust to the detection of spurious lines. But the computation time is huge, which is improper to the LDWS.

After analyzing the traditional Hough transform, this paper proposes a new lane identification method. Specifically, seed points of lane are selected firstly and then utilize Hough transform for seed points of every group. Figure 6 shows the pictures after threshold and noise filtering, represents three typical situations, the normal, existing similar interference and some deviation happens.

3.2.1. Selection and Classification of the Seed Points

As indicated in Figure 6, the lane has two main features for each scan line, because the lane has width and gray value of each side has mutation. These two features can be the criterion for selecting the lane seed points and any one possesses the features can be seen as the lane seed points. The seed points between two scanning lines can be put into one seed points group, if their pixel difference is less than a certain threshold value, otherwise they will be put into a new group. The specific steps are as follows.

Firstly, take a two-dimensional array $SeedGroup [g \times r]$ represents the x -coordinate of the seed points. Among which, g is the amount of the seed points arrays, which is also the maximum lane line numbers. According to the CCD equipment view and the highway scenario, the maximum value of g is set 4. r represents the amount of the scan lines and also the seed points y -coordinate. As the lane marks are always in the lower half of the picture, we set the value of r to a half of image height, that is, $r = 120$. This array is initialized to be 0.

Secondly, scan the image from left to right and bottom to top. When the scanning encounters a white point (x_j, y_j) , goes on with the scanning, and starts to count numbers, ends when encounters a point is not a white one. Then number of the white points is s , take the middle value of it as the seed point

$$\begin{aligned} x_s &= x_j + \frac{s}{2} \quad s \text{ is even,} \\ x_s &= x_j + \frac{(s+1)}{2} \quad s \text{ is odd.} \end{aligned} \tag{3.1}$$

Therefore, the seed point's coordinate is (x_s, y_j) . Considering the case that the lane is discontinuous which may lead to seed points separated by several scanning lines, we classify them by taking the strategy as follows. According to the CCD perspective principle, when y_j is much larger, the actual distance each pixel represents is much farther. so y_j is divided into two parts: If $y_j < 30$, we should compare the $SeedGroup$ values of all lane arrays before y_j with x_s . If it is less than a certain threshold value (when the CCD is mounted on the vehicle basically horizontal, through test the threshold is set to be 20), it classifies the seed point to this lane array. If it is larger than a certain threshold value, it indicates that a new lane is appearing. We can classify the seed point to the new lane array. Else if $y_j > 30$, we should compare the $SeedGroup$ values of the first 30 lane arrays before y_j with x_s . Classify principle is same as above.

Suppose that the lane array which the seed points are classified to is g_i . We should retain the data of lane array ($SeedGroup [g_i \times y_j], y_j$). And then continue scan starts from the 0 pixel downward, until to finish the current scanning.

Finally, continue to scan up until to complete all seed points' choice and classification. Figure 7 displays the nonzero seed points arrays $SeedGroup ([g_i \times y_j], y_j)$ ($i \in (0, g), j \in (0, r)$) for the image in Figure 6. Each array represents a lane. As can be seen, this method can accurately obtain the points of center lanes, and classify the points accurately.

3.2.2. Utilizing Hough Transform for the Seed Points of Each Group

Once the seed points have been selected and classified, and each lane can be determined utilizing Hough transform according those seed points of each group. Quantify the parameter

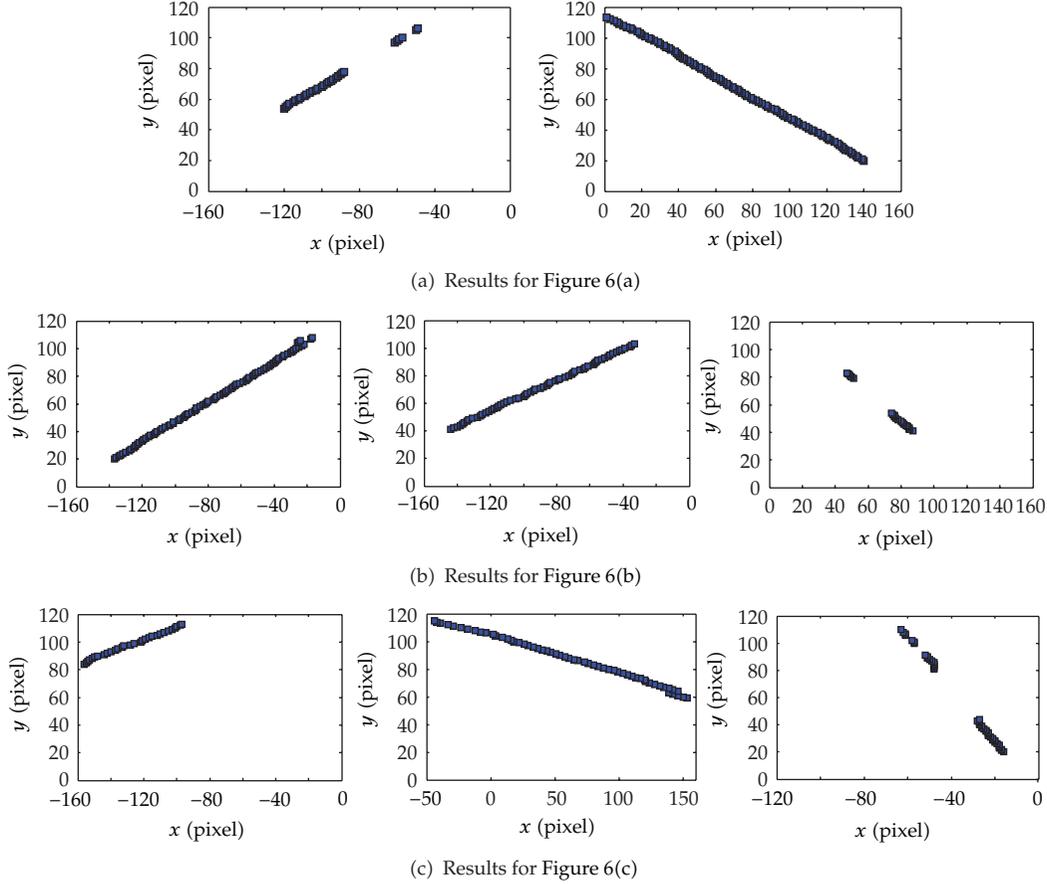


Figure 7: Seed points classification results for images in Figure 6.

space (ρ, θ) and build a two-dimensional accumulate array, then initialize the two-dimensional array at first and utilize Hough transform for the seed points of each group. We should let each point take all value on the θ axis and calculate the value of the ρ . Compare the size of the array element to get maximum value. (ρ_i, θ_j) corresponding with maximum value is the parameter corresponded with the required straight line. So we just find out the parameter (ρ_i, θ_j) of each lane.

3.2.3. Current Lane Detection with Angle Constraints

Commonly, there may exist more than one lane in the image therefore, we need to define a certain constraint to extract the current lane accurately. Because of the CCD projection transformation, the lane disappeared in a point in the image. Figure 8 displays the angular relationship diagram for multilane in the image. As shown in this figure, the angle for the left lane is $\theta_l \in (\pi/2, \pi)$ and for the right lane is $\theta_r \in (0, \pi/2)$. The farther away from the centerline of the image, the smaller is θ_l and the larger is θ_r . Therefore, we can extract the current lane in accordance with this method.

When exploiting the improved Hough transform acquiring the parameters of each straight lane, the minimum in the angle of $(0, \pi/2)$ and the maximum in $(\pi/2, \pi)$ can be

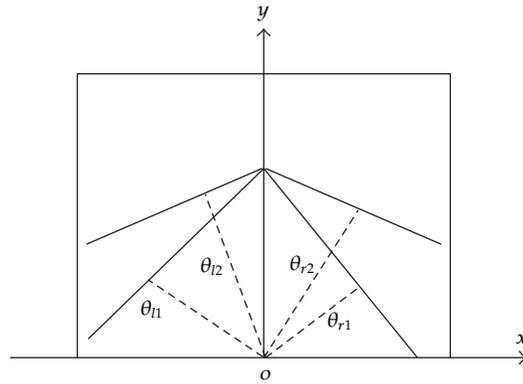


Figure 8: Diagram of angular relationship for multilane.

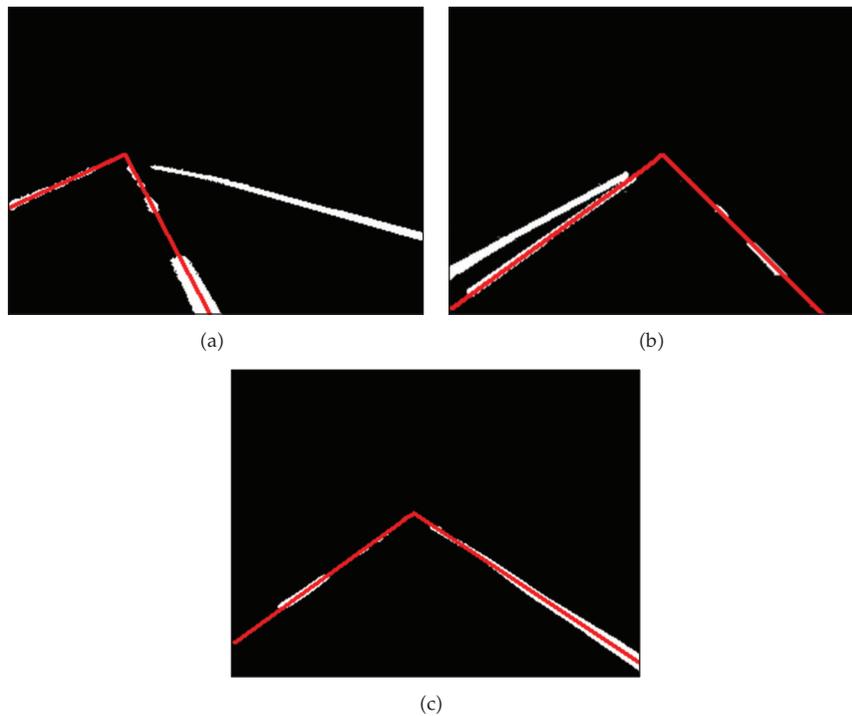
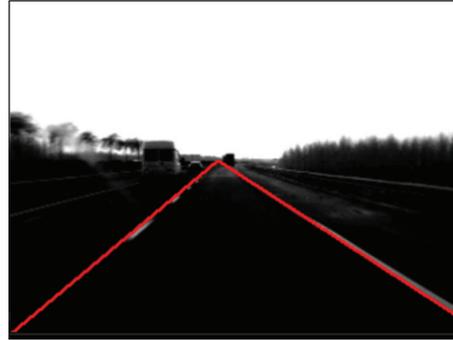
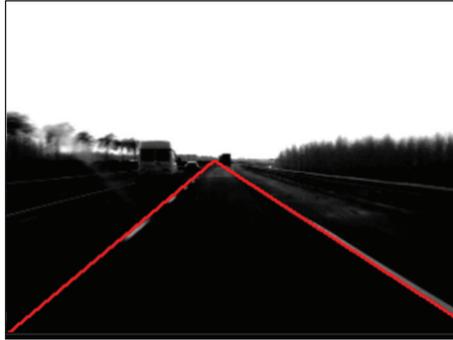


Figure 9: Detection results with angle constraints.

determined. Then the current lane can be detected according to the parameter map (ρ, θ) . Figure 9 displays the current lane detection results.

4. Experiments and Analysis

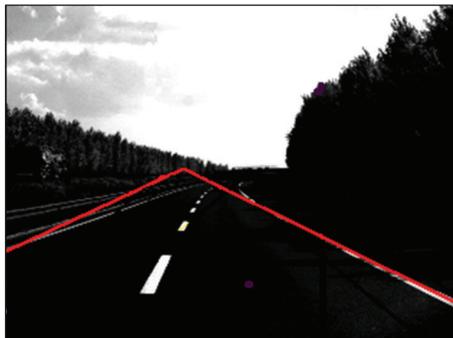
To verify the proposed lane detection method, several experiments were conducted on images and sequences on real highway scenarios with the lane departure warning system. During the experiment, the driver drives the vehicle cross the lane intentionally to see the



(a) Detection result under normal environments



(b) Detection result under strong brightness environments with interference on road



(c) Detection result when vehicle lane exchanging

Figure 10: Some typical experimental results.

system performance. The driving environment is focused on highway with different light conditions. Some typical experiments results for lane detection are shown in Figure 10. The left images are the detection results using traditional Hough transform, while the right images display the results using the improved Hough transform. Those entire images are captured after applying the intelligent regulated CCD parameters.

As can be seen from Figure 10(a), both methods can realize the lane detection under normal environments. The difference lies in the processing time of each image. The average processing time for an image using the traditional Hough transform is about 45 ms. It decreases to 25 ms when using the improved Hough transform. Moreover, the proposed lane detection method can recognize the lane correctly under abnormal environments, such as

Table 1: Statistics of our lane detection method.

Method	CCD parameters	Weather	Frames	Hit	Miss	Hit rate
Traditional Hough transform	unregulated	sunny	480	451	29	93.96%
		cloudy	256	242	14	94.53%
	regulated	sunny	734	711	23	96.87%
		cloudy	675	659	16	97.63%
Improved Hough transform	unregulated	sunny	624	607	17	97.28%
		cloudy	485	479	6	98.76%
	regulated	sunny	738	727	11	98.51%
		cloudy	626	621	5	99.20%

strong brightness. Even there are interferences on road. While the traditional method fails in this case, as displayed in Figure 10(b). It is important for the LDWS to achieve correct detection of current lane, especially when the vehicle is in process of lane exchanging. The traditional method usually takes the adjacent boundary as the current lane, which may lead to missing alarm and result in serious accident. Fortunately, the improved Hough transform is able to detection the current lane accurately, even if they are dashed lane marks, as shown in Figure 10(c) that may contribute to the importing of angle restraints.

Table 1 compares the performance of our lane detection method with traditional method. Statistic analysis also indicates that after the CCD parameters regulating, the performances for both of the traditional and improved Hough transform are enhanced. While the proposed Hough transform presents the best performance. As to the weather conditions, cloudy day shows better than sunny day for the reason that the road surface is more dark, which is helpful to enlarge the contrast between the surface and the white lane marks.

5. Conclusions

In this paper, a robust lane detection method based on intelligent CCD parameters regulation by combining the improved Hough transform with certain angle restraints is presented. The main contributions of this paper are as follows.

- (1) The CCD parameters, such as the brightness, gain, and exposure time, are regulated based on the information in the specialized feature regions. Test results indicate that the adjusting of CCD parameters can enhance the contrast between the lane marks and road surfaces and reduce background noise, which is helpful to improve the real time of the algorithm.
- (2) An improved Hough transform algorithm is proposed after analyzing the advantages and disadvantages of traditional one, which optimizes the selection of seeds and introducing the angle constraints.
- (3) Several road experiments were conducted to validate the proposed method. Compared with the traditional Hough transform, results show that the improved method can detect the current lane with lower processing time and higher detection accuracy.

In this case, the lane model is taken as straight line and ignoring the curve line model that really existing in real environments. In the future work, this case should be taking into

account for the lane departure warning system. Moreover, the vehicle steering controlling is necessary according to the lane detection results.

Acknowledgments

This research was financed by the National Natural Science Foundation of China (61104165) and the Fundamental Research Funds for the Central Universities (DUT12JR08 and 20116217).

References

- [1] R. Sánchez-Mangas, A. García-Ferrrer, A. de Juan, and A. M. Arroyo, "The probability of death in road traffic accidents. How important is a quick medical response?" *Accident Analysis and Prevention*, vol. 42, no. 4, pp. 1048–1056, 2010.
- [2] W. Zhang, O. Tsimhoni, M. Sivak, and M. J. Flannagan, "Road safety in China: analysis of current challenges," *Journal of Safety Research*, vol. 41, no. 1, pp. 25–30, 2010.
- [3] W. H. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, China Communications Press, Beijing, 2012.
- [4] W. H. Wang, H. W. Guo, H. W. Bubb, and K. Ikeuchi, "Numerical simulation and analysis procedure for model-based digital driving dependability in intelligent transport system," *KSCE Journal of Civil Engineering*, vol. 15, no. 5, pp. 891–898, 2011.
- [5] World Health Organization, *World Health Statistics 2008*, World Health Organization, Geneva, Switzerland, 2008.
- [6] J. H. Zhou, J. Qiu, X. C. Zhao et al., "Road crash in China from 2003 to 2005," *Chinese Journal of Traumatology*, vol. 11, no. 1, pp. 3–7, 2008.
- [7] M. Karacasu and A. Er, "An analysis on distribution of traffic faults in accidents, based on driver's age and gender: Eskisehir case," *Procedia—Social and Behavioral Sciences*, vol. 20, pp. 776–785, 2011.
- [8] M. G. Masuri, K. A. M. Isa, and M. P. M. Tahir, "Children, youth and road environment: road traffic accident," *Procedia—Social and Behavioral Sciences*, vol. 38, pp. 213–218, 2012.
- [9] A. Spoerri, M. Egger, and E. von Elm, "Mortality from road traffic accidents in Switzerland: longitudinal and spatial analyses," *Accident Analysis and Prevention*, vol. 43, no. 1, pp. 40–48, 2011.
- [10] A. Bener, M. G. A. Al Maadid, T. Özkan, D. A. E. Al-Bast, K. N. Diyab, and T. Lajunen, "The impact of four-wheel drive on risky driver behaviours and road traffic accidents," *Transportation Research F*, vol. 11, no. 5, pp. 324–333, 2008.
- [11] P. Angkititrakul, R. Terashima, and T. Wakita, "On the use of stochastic driver behavior model in lane departure warning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 174–183, 2011.
- [12] N. Minoiu Enache, M. Netto, S. Mammar, and B. Lusetti, "Driver steering assistance for lane departure avoidance," *Control Engineering Practice*, vol. 17, no. 6, pp. 642–651, 2009.
- [13] E. A. Schmidt, M. Schrauf, M. Simon, A. Buchner, and W. E. Kincses, "The short-term effect of verbally assessing drivers' state on vigilance indices during monotonous daytime driving," *Transportation Research F*, vol. 14, no. 3, pp. 251–260, 2011.
- [14] J. C. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 20–37, 2006.
- [15] Đ. Obradović, Z. Konjović, E. Pap, and I. J. Rudas, "Linear fuzzy space based road lane model and detection," *Knowledge-Based Systems*. In press.
- [16] Y. Wang, N. Dahnoun, and A. Achim, "A novel system for robust lane detection and tracking," *Signal Processing*, vol. 92, no. 2, pp. 319–334, 2012.
- [17] S. Nedeveschi, R. Danescu, T. Marita et al., "A sensor for urban driving assistance systems based on dense stereovision," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV'07)*, pp. 276–283, Istanbul, Turkey, June 2007.
- [18] J. G. Wang, C. J. Lin, and S. M. Chen, "Applying fuzzy method to vision-based lane detection and departure warning system," *Expert Systems with Applications*, vol. 37, no. 1, pp. 113–126, 2010.

- [19] B. F. Wu, C. J. Chen, Y. P. Hsu, and M. W. Chung, "A DSP-based lane departure warning system," in *Proceedings of the 8th WSEAS International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering*, Bucharest, Romania, 2006.
- [20] E. Shang, J. Li, X. J. An, and H. H. He, "A real-time lane departure warning system based on FPGA," in *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems*, pp. 1243–1248, Washington, DC, USA, 2011.
- [21] P. Y. Hsiao, C. W. Yeh, S. S. Huang, and L. C. Fu, "A portable vision-based real-time lane departure warning system: day and night," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 2089–2094, 2009.
- [22] L. S. Jin, R. B. Wang, T. H. Yu, L. Guo, and L. Yang, "Lane departure warning system based on monocular vision," in *Proceedings of the 14th World Congress on Intelligent Transportation Systems*, Beijing, China, 2007.
- [23] S. Suh and Y. Kang, "A robust lane recognition technique for vision-based navigation with a multiple clue-based filtration algorithm," *International Journal of Control, Automation and Systems*, vol. 9, no. 2, pp. 348–357, 2011.
- [24] Z. W. Kim, "Robust lane detection and tracking in challenging scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 16–26, 2008.
- [25] V. Shapiro, "Accuracy of the straight line Hough Transform: the non-voting approach," *Computer Vision and Image Understanding*, vol. 103, no. 1, pp. 1–21, 2006.
- [26] A. L. Fisher and P. T. Highnam, "Computing the Hough transform on a scan line array processor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 3, pp. 262–265, 1989.
- [27] M. C. K. Yang, J. S. Lee, C. C. Lien, and C. L. Huang, "Hough transform modified by line connectivity and line thickness," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 8, pp. 905–910, 1997.
- [28] L. A. F. Fernandes and M. M. Oliveira, "Real-time line detection through an improved Hough transform voting scheme," *Pattern Recognition*, vol. 41, no. 1, pp. 299–314, 2008.

Research Article

A Cloud-Computing-Based Data Placement Strategy in High-Speed Railway

Hanning Wang,¹ Weixiang Xu,² Futian Wang,¹ and Chaolong Jia¹

¹ State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

² School of Traffic and Transportation, Beijing JiaoTong University, Beijing 100044, China

Correspondence should be addressed to Hanning Wang, 06114158@bjtu.edu.cn

Received 28 August 2012; Revised 31 October 2012; Accepted 21 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Hanning 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.

As an important component of China's transportation data sharing system, high-speed railway data sharing is a typical application of data-intensive computing. Currently, most high-speed railway data is shared in cloud computing environment. Thus, there is an urgent need for an effective cloud-computing-based data placement strategy in high-speed railway. In this paper, a new data placement strategy named hierarchical structure data placement strategy is proposed. The proposed method combines the semidefinite programming algorithm with the dynamic interval mapping algorithm. The semi-definite programming algorithm is suitable for the placement of files with various replications, ensuring that different replications of a file are placed on different storage devices, while the dynamic interval mapping algorithm ensures better self-adaptability of the data storage system. A hierarchical data placement strategy is proposed for large-scale networks. In this paper, a new theoretical analysis is provided, which is put in comparison with several other previous data placement approaches, showing the efficacy of the new analysis in several experiments.

1. Introduction

With the development and popularity of information technology, internet is gradually growing into various computing platforms. Cloud computing is a typical network computing mode, which emphasizes the scalability and availability of running large-scale application in virtual computing environment [1]. A large-scale network application based on cloud computing demonstrates features of distribution, heterogeneousness, and the trend of data intensity [2]. In cloud computing environment, data storage and operation is provided as a service [3]. There are various types of data, including common files, large binary files such as virtual machine image file, formatted data like XML, and relational data in database. Thus, a

distributed storage service of cloud computing has to take large-scale storage mechanism for various data types into account, as well as the performance, reliability, security, and simplicity of data operation. As an important component of China's transportation science data sharing system, high-speed railway data is the key to optimizing the operating organization. High-speed railway data sharing system has the characteristics of typical data-intensive application [4–6], to which the management of a large amount of distributed data is crucial. It is mainly reflected in the fact that the data size it processes is often up to TB or even PB level, which includes both existing input data source and the intermediate/final result data produced in the process.

In the process of implementing and executing data-intensive applications under the environment of cloud computing, as well as the process of establishing a large-scale storage system to meet the demand of a fast growing data storage volume, the main challenge is how to effectively distribute data at Petabyte level to hundreds of thousands of storage devices. Thus, an effective data placement algorithm is needed.

2. Goal of Designing High-Speed Railway Data Placement Strategy

The network storage system under cloud computing environment consists of thousands, and even ten thousands of storage devices. Different systems have different underlying devices, for example, the storage device set could be chunk device disk for SAN and GFPS, or OSD (object storage device) for object storage systems Lustre and ActiveScale, or PC for PVFS and P2P [7]. A data placement strategy mainly solves the problem of selecting storage device for data storage. An effective mechanism shall be adopted to establish the mapping relationship between data sets and storage device sets. Then, data sets generated by applications in the storage system are placed into different storage device sets. Meanwhile, certain particular goals need to be met, and different data placement strategies are designed for different purposes. For example, the stripping technology in RAID is mainly designed to acquire aggregated I/O bandwidth. The strategy of placing several replications of the data into different devices is mainly for the purpose of fault tolerance and data reliability improvement. Distributing data equally could realize a more balanced I/O load.

The high-speed railway data placement strategy under cloud computing environment is designed to meet the following goals.

2.1. Fairness

The size of the data stored in each device is proportional to the storage volume of that device [8].

2.2. Self-Adaptability

With the elapse of time, the volume of storage devices is dynamic and varied. Take the case of adding a new device and the case of deleting an existing device for example. When the scale of the storage system changes, a data placement strategy is applied to reorganize the data, making the data distributed to device sets satisfy the fairness criteria over again. Furthermore, it needs to be ensured that the migrated data volume is close to the optimal migration data volume. This would reduce the overheads of data migration. The optimal data volume to

migrate is equal to the data volume that is acquired by the added device, or equal to the data volume on the deleted device. The self-adaptability of the data placement strategy is measured by the ratio of its actual migrated data volume to the optimal migration data volume. Therefore, the ratio value of 1.0 represents the optimal condition.

2.3. Redundancy

Getting several replications copied for the data, or enabling the data remain accessible through the use of erasure code when one of the replications is lost. So that fairness can balance the IO loads, self-adaptability can reensure fairness in accordance with storage scale change, and the data size migrated and the IO bandwidth occupied can also be decreased. Finally, the data reliability can be improved.

2.4. Availability

It is crucial that a system could be normally accessed in all cases. Once the system is unavailable, all functions would fail to perform normally. To improve system availability, it is necessary to regularly have the data location adjusted according the availability of storage devices, thus maximizing the system's availability [9].

2.5. Reliability

It indicates whether the system could be normally accessed during a certain period of time. As the large-scale storage system contains thousands of storage devices, the probability of disk failure is rather high. When applying a data placement strategy, indicators of reliability such as data size need to be used in designing parameters of the placement strategy. Thus, a storage system with higher reliability is obtained.

2.6. Space-Time Effectiveness

It means that few time and space is used in calculating data location along with the data placement strategy.

When designing the data placement strategy for large-scale network storage system, certain particular goals need to be met depending on different application demands. However, it is impossible to meet all goals at the same time.

3. Related Work

Some data management systems under the cloud computing environment have already been emerged currently, for example, Google File System [10] and Hadoop [11, 12], both of which have hidden the infrastructure used to store application data from the user. Google File System is mainly used for Web search application, but not for process application under the cloud computing environment. Hadoop is a more commonly distributed file system, which is used by many companies including Amazon [13] and Facebook. When Hadoop file system receives a file, the system will automatically separate the file into several chunks,

each of which is randomly placed into a cluster. Cumulus project [14] has proposed a cloud architecture of single data center environment. However, the above-mentioned cloud data management systems did not study the data placement problem of data-intensive process applications under the cloud environment. Finally, let us look into several examples of existing popular large-scale data storage systems. Commercial Parallel File System (CPFS) [15, 16] divides a file into data chunks of the same size, which are stored on different disks of the file system in the form of rotation. Parallel Virtual File System (PVFS) [17] with open-source codes divides the file into strip and chunk and adopts the method of placing sliced data on multiple IO nodes in rotation. The data slice size of PVFS is a constant. PVFS data does not have any fault tolerance function. Panasas [18] is an object-oriented file system, where the data are allocated to underlying smart object storage device (OSD) in the unit of object [19]. A file is divided into strips, and each strip unit is stored on multiple OSD in the form of object. Upon initial placement, objects are fairly distributed between OSD devices using the random method.

PanFS, developed by the Panasas Company, is a Linux cluster file system based on object storage [20]. It is the core part of ActiveScale storage system. At first, these file systems divide the file into strips, and then allocate each strip to the underlying smart OSD in unit of object. The distribution of files across multiple OSDs is realized based on the round-robin algorithm. The size of the data object is random, and it could increase accordingly with the increase of file size without modifying the metadata mapping chart on the metadata server.

The object-oriented file system Lustre is a transparent global file system. The Lustre file system treats the file as an object that is located by the metadata server, which then directs the actual file I/O request to the corresponding object storage targets (OSTs). Since a technology is adopted where the metadata is separated from the storage data, the computing resources could be fully separated from storage resources [21]. Thus, the client could focus on the I/O request from users and applications. Meanwhile, the OST and metadata server could focus on data reading, transmitting, and writing.

All storage nodes in the COSMOS parallel file system [22] are divided into several strips. Each COSMOS file is stored in a certain strip. And the length and logic chunk length of the strip are related to the disk speed and file access mode of applications. This type of data placement strategy has features such as high performance, large file suitability, and high degree of parallelism. Through the stripped subfile, COSMOS is directly saved on local disks in the form of common JFS file. Thus, the expression direct management of disks is avoided while increases the overheads when entering into the core of VFS/Vnode for the second time.

4. Study and Analysis of Existing Data Placement Strategy

Here are some currently popular data placement algorithms. Standard hashing is the simplest homogeneous (indicating that all storage devices have the same volume) placement algorithm, which can ensure fairness. But when the storage scale varies, the locations of all the data have to be changed as well.

Consistent Hashing [23] uses the *hash* function to map a device to the continuum, and then the hash function $h_2(x)$ is used to evenly map the data to that continuum. Then data is allocated to that device represented by the node which is nearest to the data itself. Since devices are not evenly distributed on the continuum, each device is virtualized to $k \log |N|$ devices (where k is a constant) to ensure the fairness of data allocation. The data size of this device is equal to the total size of data allocated to virtual nodes. When a storage device is

added to the system, only parts of the data located on the left and right neighbor nodes are to be migrated to that device. Consistent Hashing has a high degree of self-adaptability, and this mechanism takes up a space of $O(n^2 \log n)$.

As a matter of fact, the data storage under the cloud computing environment is heterogeneous, which means that there is great volume discrepancy between storage devices. Therefore, the consistent Hashing algorithm is improved as follows: the virtual nodes on the continuum are allocated based on the weight of a device. The device with greater weight covers more virtual nodes on the continuum. However, this approach would introduce large amount of virtual nodes in a heterogeneous storage system with extremely significant weight discrepancy, and this would increase the space complexity of the algorithm.

In order to solve the problem of space waste with consistent hashing, a segmentation method based on the unit interval is brought forward. In this method, the interval is divided into unit subintervals with identical length, and each device occupies an interval. When a device is added, part of the data on other devices is migrated to this new device. When a device is deleted, the data on the last device is equally migrated to remaining devices, and the data on the device to be deleted is migrated to the last device, and then the very device is finally deleted. In this way, the fairness could be guaranteed. During the device addition, the data migration volume is 1 time the optimal data migration volume. During the device deletion, the data migration volume is 2 times the optimal data migration volume. $O(\log n)$ steps are needed to locate a specific data, which takes longer time than locating data with consistent hashing, but only a space of $O(n \log n)$ digit is occupied here. Compared to the consistent hashing, this algorithm exchanges the time for space. It is not suitable for a storage system that has a demanding requirement for rapid data search. Furthermore, this algorithm's self-adaptability is not as high as consistent hashing.

In order to solve the problem of space waste resulting from the consistent hashing's introduction of virtual nodes, the linear method and logarithm method are proposed. In the linear method, the weight for a device is introduced similarly. Suppose w_i indicates the weight for device i , and $d_i(x)$ indicates the distance between the hash values of device i and data x . The linear method would select the device, which has the lowest value of $H_i = d_i(x)/w_i$, to store data x .

As the storage scale changes, the linear method could guarantee that data would only be migrated between the added/deleted device and other devices. There would be no data migration between other devices. The logarithm seeks to find a device that brings the the minimum value for the function $H_i = -(\ln(1 - d_i(x)))/w_i$. In the absence of the virtual nodes, the logarithm performs with better fairness than linear one, but it would take a longer time to locate data.

As a result, a data object placement algorithm based on the dynamic interval mapping is proposed [22]. The unit space is divided into multiple subintervals according to the device's weight. And then a mapping relationship between the device and subinterval is established. Based on the interval where the data falls in, data is allocated to the device corresponding to that interval. This approach has better fairness and self-adaptability, where the time consumption in locating data will increase with the expansion of the number of storage devices. But if the number of storage devices is extremely large, when a device is added or deleted, the system is required to communicate with all other storage devices for data migration, which will bring tremendous overheads. Furthermore, the time consumption in locating data will increase with the expansion of the number of storage devices as well.

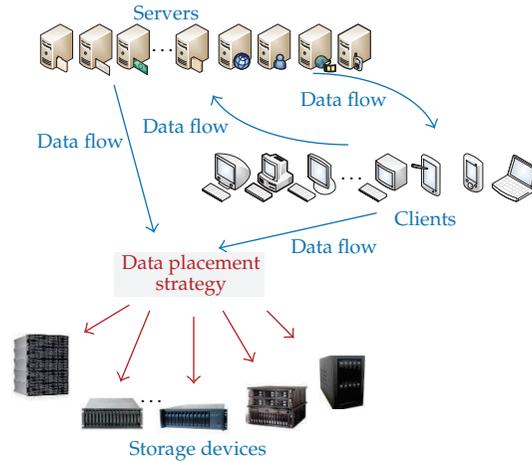


Figure 1: Direct management.

5. A Hierarchical Structure Based on Cloud Computing

With the expansion of network scale, the number of data storage devices keeps increasing. The existing data placement algorithm is insufficient to address the system's self-adaptability. Adding new or removing existing devices could lead to a new data placement over again, which will result in an increase of data migration overheads so that the occupation of IO bandwidth is inevitable [24, 25]. Therefore, the data reliability cannot be guaranteed, and the overheads are too large to use a duplicate copy for data reliability assurance [26]. Thus, a data placement strategy based on a hierarchical structure is brought forward in this paper for the purpose of making up for the shortage of existing data placement algorithm, addressing the system's self-adaptability, guaranteeing data reliability, and improving the efficiency of data access.

In the proposed approach, each single storage device is directly managed through a common data placement strategy, as shown in Figure 1.

The hierarchical structure could reduce the time consumption of data query and location. As a result, a data placement strategy of hierarchical structure is more suitable for data management under cloud computing environment, as shown in Figure 2.

It is assumed in this paper that large amount of storage devices are heterogeneous in the storage system under cloud computing environment. That is to say, the storage volume of every single device is different from one another. These storage devices are grouped into several device sets that count relatively less in number. When storing the file data, it is first located on a device set, and then the file data is stored inside the device set. So that the locality of file data within this device set is ensured, and this helps improve the speed for data reading and writing.

In the case of data placement for file with several replications, different replications of the same file should be placed onto different device sets as many as possible. So that when a certain storage device within a storage device set cannot function properly, the client could get the target file data located on other device sets as usual. Thus, it could improve the availability and reliability of files.

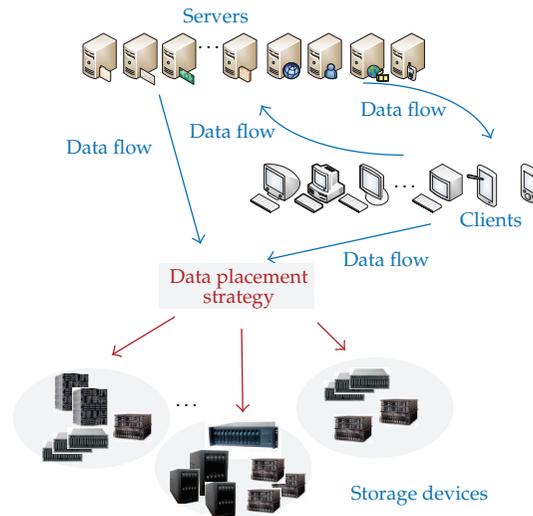


Figure 2: Hierarchical structure.

In the data placement strategy of hierarchical structure, when a storage device is to be added, it is designed to allocate the newly added storage device to a device set; when a storage device is to be deleted from a device set, the migration data could be constrained to other different storage devices within that device set. This would reduce the overheads in communication with large amount of storage devices within other storage device sets. The I/O bandwidth occupation would be reduced during data migration. When an aged storage device needs to be replaced with a new one, firstly the data on the original device is transferred to the new device. Since the new storage device in replacement outperforms the original one in both the storage volume and read/write performance, fairness is disrupted between each storage device within the device set in regard of data storage. Therefore, data is migrated between the new storage device and other ones within that device set in order to meet fairness criteria between each storage device within that set.

6. Algorithm Description

We would group larger amount of heterogeneous storage devices into less amount of device sets. The number of already grouped sets is to be kept unchanged. The total storage volume of different storage device sets should remain the same. Files and their various numbers of duplicate copies are to be mapped to different device sets for storage using an algorithm based on semidefinite programming. Files are sliced within the device set, and then the data slices are mapped to devices with different volumes in the set using a dynamic interval mapping method.

6.1. The Semidefinite Programming Algorithm

So that the problem of data copies placement is converted into a problem of seeking semidefinite programming, different copies of a file are placed on different storage device sets. Meanwhile, according to the algorithm, the file is located on a device set and stored into

Formal description of a semi-definite programming problem

Solution:

$$\min \sum_{i \neq j}^n C(i, j) \cdot l_i \cdot l_j$$

Satisfying:

l_i, l_j are unit vectors; $l_{i,i} = 1$
 l_i and l_j form the matrix $L = [l_i \cdot l_j]$ and all its
 eigenvalues are greater than or equal to 0,
 that is, matrix L is semi-definite.

Algorithm 1: Formal description.

various devices within the set strip by strip; thus, the file locality is ensured. The file data could be immediately accessed upon one time of locating, so that the file access speed is improved.

Function $C(i, j) = (i, j) = 1$ is right only when i and j are two different copies of the same file, or when j stands for the copy of i . If not, $C(i, j) = 0$. Also, $C(i, j) = 0$ when i equals j . An associated matrix C is constructed using $C(i, j)$. C can represent the relationship between all files, that is, which file owns and which file copies. The Algorithm 1 converts the problem of data copy placement into the formalized description of a semidefinite programming problem.

Solution to the semidefinite programming problem can produce a semidefinite matrix $L = [l_i \cdot l_j]$. And further processing of the semidefinite matrix L can obtain the device sets, where each file copy is stored in the storage system.

6.2. Dynamic Interval Mapping Algorithm

Supposing some device set D contains n devices, that is, d_1, d_2, \dots, d_n . All these n devices have different volumes, respectively, D_1, D_2, \dots, D_n , so that the ratio of the weight of each device volume to the total volume within this device set is $r_i = D_i / (D_1 + D_2 + \dots + D_n)$, where $i = 1, 2, \dots, n$ and $r_0 = 0$. It is known that $\sum_{i=1}^n r_i = 1$. So then we segment a subinterval with the length of $w_i = [r_{i-1}, r_{i-1} + r_i]$ for each device D_i in the interval $[0, 1]$. When the file is allocated into a device set, it is divided into data chunk sets $S : s_1, s_2, \dots, s_m$ with the same size, and, then m data chunks are mapped to devices with different weight values in the set (Algorithm 2).

The hash function $h(S) : \rightarrow [0, 1]$ is used to map the data chunks to the interval $[0, 1]$. If $h(s_i) \in w_i$, then the data chunk s_i is allocated to the device mapped by the interval w_i .

7. Experiment and Analysis

In this paper, the two key algorithms in hierarchical data placement, namely, Semidefinite programming (SDP) algorithm and dynamic interval mapping algorithm, are implemented on Matlab platform. Matrix is the basic unit of Matlab language, which could be directly used in matrix calculation. Therefore, Matlab could be directly applied to solve complicated problems such as optimization or linear programming. The semidefinite programming problem we need to solve in this paper will be described in a mathematic formalized

```

Pseudocode of the algorithm
Initialization:
  Device set  $D$ , Data set  $S$ , and subinterval set  $w$ ;
Input: data chunk  $s_i$ 
Program main:  $k = h(s_i)$ ;
  for ( $j = 1; j < n; j ++$ )
    if ( $k \in w_j$ )
      Placing the data chunk  $s_i$  on the device  $d_j$ 
Output: data volume stored on the device  $d_j$ 

```

Algorithm 2: Formal description.

matrix. Furthermore, it is easy to formalize a dynamic interval mapping problem into a formalized matrix, which is suitable to implement in Matlab environment. Meanwhile, Matlab features an abundant toolbox and module set. In order to seek a solution to the semidefinite programming problem, a toolbox that provides support for Matlab to solve SDP problem should be installed.

7.1. Fairness Analysis on Semidefinite Programming Algorithm

Suppose that each file has 5 copies. Then, respectively, distribute 100, 200, 300, and 400 files into 10 device sets and 20 device sets using the semidefinite programming method. The deployment is shown in Figures 3 and 4. The experiment has shown that files can be fairly evenly distributed to multiple device sets using the semidefinite programming. It has been illustrated that this approach could ensure the fairness of file data layout.

7.2. Reliability Analysis on Semidefinite Programming Algorithm

Now let us further discuss the situation for placement of 5 copies of the same file, that is, the problem regarding whether all the 5 copies of the same file have been placed into different device sets. As shown in Table 1, when 400 files (with 2,000 copies) are distributed to 10 and 20 device sets, all the 5 copies of, respectively, 299 and 372 files are completely distributed to 5 different device sets. Other files which do not include these copies failed to do this. There are 2 of 5 copies of one file that are allocated to the same device set. As a result, the semidefinite programming algorithm has shown a better performance to allocate different copies of a file into different storage device sets. Thus, the probability of data loss due to device failure is reduced, and the data reliability is improved. Based on the principle of random function, it can be inferred that the probability for data allocated to each subinterval using the dynamic interval mapping algorithm is proportional to the length of each interval. Similarly, the data volume of all the devices inside the device set is proportional to its overheads. It has been proved that when the storage nodes inside the storage device set change, the dynamic interval mapping method can minimize the overheads of data migration under the condition that the number of storage nodes is not extremely high. So it eliminates the overheads of communicating and migrating data caused by the change of the number of storage nodes, when directly managing a very large amount of storage devices. As a new device is added, the subinterval occupied by each device within a device

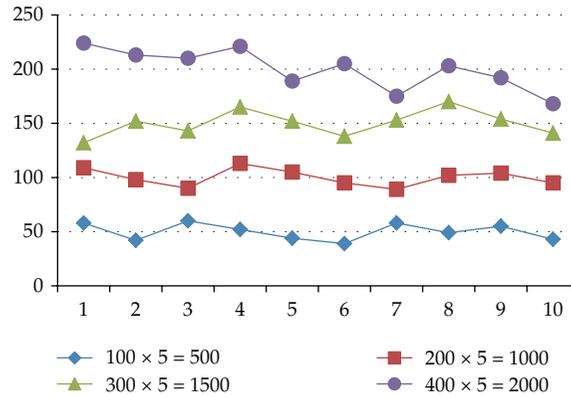


Figure 3: Files distribution over 10 device sets.

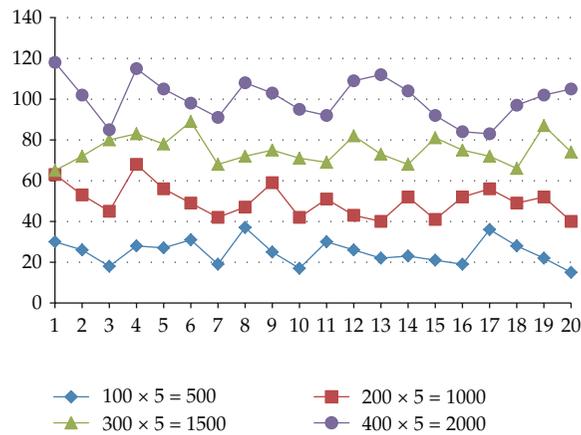


Figure 4: Files distribution over 10 device sets.

set changes correspondingly, reallocating the interval occupied by existing devices and the corresponding data chunk to the new device, in order to realize fairness over again. The overheads of communicating and transferring data are constrained to only those few devices inside the device set.

7.3. Fairness Analysis on Dynamic Interval Mapping Algorithm

Firstly, the fairness of dynamic interval mapping algorithm is tested. Let us take a look at the file data volume stored on each storage device within a device set. When 1,000 files are stored in 10 device sets, 100 files are stored in the no. 5 device set as indicated in Figure 3. Then, we assume that there are 10 storage devices inside the no. 5 device set. And the 100 files are stripped into 1,500 data strips, which are stored onto the 10 storage devices by means of dynamic interval mapping algorithm. For all these 10 storage devices, the percentage of each device's storage volume in their total storage volume, as well as the interval length (w_j) corresponding to that percentage are all shown in Table 2.

Table 1: Distribution situation of replications.

Replications distribution	10 device sets	20 device sets
100 files (500 replications)	79	93
200 files (1000 replications)	154	186
300 files (1500 replications)	237	266
400 files (2000 replications)	299	372

Table 2: Storage percentage of each device relative to the total storage volume.

Device code	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10
%	5	10	16	11	20	9	8	10	6	5
Interval	0	0.05	0.15	0.31	0.42	0.62	0.71	0.79	0.89	0.95
	~	~	~	~	~	~	~	~	~	~
	0.05	0.15	0.31	0.42	0.62	0.71	0.79	0.89	0.95	1

Based on the dynamic interval mapping algorithm and above-mentioned volume of each storage device, the stripped 1,500 data strips are equally stored into these 10 storage devices. The theoretical allocation situation is shown in Table 3.

When implementing the dynamic interval mapping algorithm, the hash function $h(s_i)$ is used to map data chunk s_i to a random number between $(0, 1)$. If $h(s_i) \in w_j$, $(0 < j \leq 10)$, the data chunk s_i is placed into the storage device j . Consequently, all the 1,500 data strips are stored into the 10 storage devices. Comparison between the actual allocation situation and theoretical situation is shown in Figure 5.

7.4. Self-Adaptability Analysis on Dynamic Interval Mapping Algorithm

Let us test the self-adaptability of dynamic interval mapping algorithm. The cases of removing a storage device and adding a new storage device are, respectively, considered.

7.4.1. Removing a Storage Device

Let us examine the file data volume migrated between other storage devices when a storage device is deleted from a device set. For example, Table 4 shows the situation when no. 7 device is deleted from the device set. The percentage of each remaining device's storage volume, and the interval length corresponding to that percentage are shown in the Table 4.

When removing the no. 7 storage device, the data on that device is migrated to the remaining 9 storage devices. The situation of change relating to actual data migration is shown in Figure 6.

From Figure 7 above, we can see that after deletion of no. 7 device, those remaining storage devices in the device set can still store data according to the percentage of each one's storage volume in the total remaining storage volume.

Table 3: Theoretical data allocation on 10 storage devices.

Device code	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10
Data allocation	75	150	240	165	300	135	120	150	90	75

Table 4

Device code	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10
%	5.4	10.9	17.4	12	21.7	9.8	0	10.9	6.5	5.4
	0	0.054	0.163	0.337	0.457	0.674	0	0.797	0.881	0.946
Interval	~	~	~	~	~	~	~	~	~	~
	0.054	0.163	0.337	0.457	0.674	0.772	0	0.881	0.946	1

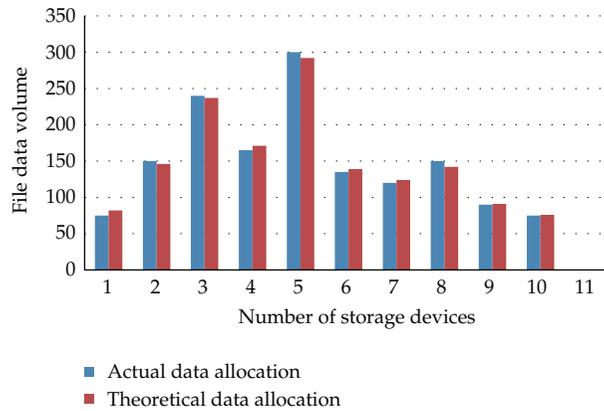


Figure 5: Comparison between actual and theoretical allocation.

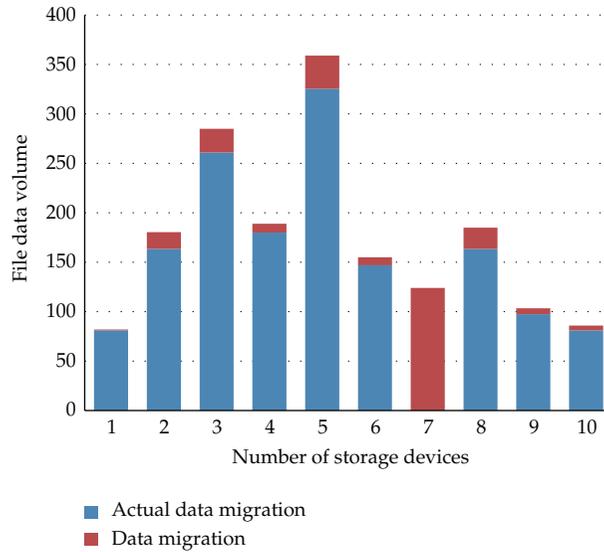


Figure 6: Data migration after deleting no. 7 device.

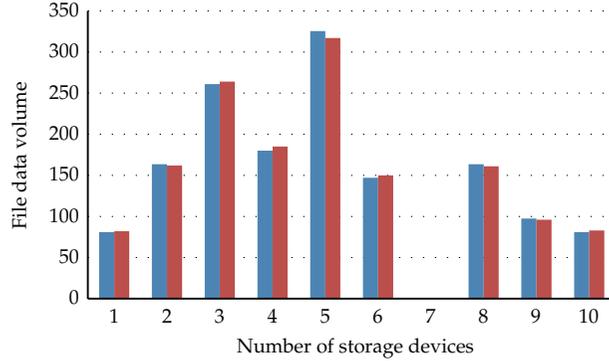


Figure 7: Comparison between the actual and theoretical allocation after deleting no. 7 device.

7.4.2. Adding a New Storage Device

Now let us examine the situation when a storage device is added to that device set. The case is similar to the above-mentioned situation when a storage device is removed. We would follow these steps below.

- (1) First, when a new storage device is added to that device set, the percentage of each device relative to the total storage volume is recalculated. And the interval length (w_j) corresponding to that percentage is redefined as well.
- (2) Then the difference between the original interval length and the revised one following the addition of a storage device is calculated. And the data corresponding to that length difference is what to be migrated into the new storage device.
- (3) Dynamic interval mapping algorithm is used to place the migrated data into the newly added storage device. The Hash function $h(s_i)$ is used to map the data strip s_i to a random number between $(0, 1)$. If $h(s_i) \in w_j, (0 < j \leq 10)$, the data chunk s_i is placed onto device j .

8. Conclusion

A hierarchical structure data placement algorithm under the cloud computing environment is proposed in this paper. The proposed algorithm combines the semidefinite programming algorithm with the dynamic interval mapping method. The semidefinite programming method would distribute the data of a file with replications to grouped device sets. Experiments have demonstrated that this method could guarantee the data reliability and high-speed file accessibility. The dynamic interval mapping method could distribute data fairly to devices with different volumes inside the device set. The self-adaptability of this method is proved theoretically.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (General Projects) (Grant no. 61272029), National Key Technology R&D Program (Grant No.: 2009BAG12A10), China Railway Ministry Major Program (2008G017-A), and State Key

Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China (Contract no. RCS2009ZT007).

References

- [1] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," *Journal of Internet Services and Applications*, vol. 1, no. 1, pp. 7–18, 2010.
- [2] C. Miceli, M. Miceli, S. Jha, H. Kaiser, and A. Merzky, "Programming abstractions for data intensive computing on clouds and grids," in *Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'09)*, pp. 478–483, Shanghai, China, May 2009.
- [3] K. Liua and L. J. Dong, "Research on cloud data storage technology and its architecture implementation," *Procedia Engineering*, vol. 29, pp. 133–137, 2012.
- [4] J. Ma, "Managing metadata for digital projects," *Library Collections, Acquisition and Technical Services*, vol. 30, no. 1-2, pp. 3–17, 2006.
- [5] 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.
- [6] W. Wang, *Vehicle's Man-Machine Interaction Safety and Driver ASSiStance*, China Communications Press, Beijing, China, 2012.
- [7] V. T. Tran, G. Antoniu, B. Nicolae, L. Bougé, and O. Tatebe, "Towards a grid file system based on a large-scale BLOB management service," in *Grids, P2P and Services Computing*, pp. 7–19, 2010.
- [8] L. Amsaleg, M. J. Franklin, A. Tomasic, and T. Urhan, "Improving responsiveness for wide-area data access," *Data Engineering*, vol. 20, pp. 3–11, 1997.
- [9] A. Deshpande and Z. Ives, "Adaptive query processing," *Foundations and Trends in Databases*, vol. 1, no. 1, pp. 1–140, 2007.
- [10] S. Ghemawat, H. Gobioff, and S. T. Leung, "The google file system," in *Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP'03)*, pp. 29–43, ACM, New York, NY, USA, October 2003.
- [11] Apache Hadoop, <http://hadoop.apache.org/>.
- [12] "Hadoop distributed file system," http://hadoop.apache.org/docs/r0.18.0/hdfs_design.pdf.
- [13] Amazon Elastic MapReduce, <http://aws.amazon.com/elasticmapreduce/>.
- [14] J. Tao, M. Kunze, A. C. Castellanos, L. Wang, D. Kramer, and W. Karl, "Scientific cloud computing: early definition and experience," in *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC'08)*, pp. 825–830, Dalian, China, September 2008.
- [15] Y. Zhai, M. Liu, J. Zhai, and X. Ma, "Cloud versus in-house cluster: evaluating Amazon cluster compute instances for running MPI applications," in *Proceedings of the SC'11 State of the Practice Reports*, 2011.
- [16] C. H. C. Evangelinos, "Cloud computing for parallel scientific HPC applications: feasibility of running coupled atmosphere-ocean climate models on Amazon's EC2," in *Proceedings of the 1st Workshop on Cloud Computing and Its Applications (CCA'08)*, October 2008.
- [17] P. H. Carns, I. W. Ligon, R. Ross, and R. Thakur, "PVFS: a parallel file system for linux clusters," in *Proceedings of the 4th Annual Linux Showcase and Conference*, Atlanta, Ga, USA, 2000.
- [18] "Cloud computing with parallel storage," <http://www.panasas.com/blog/cloud-computing-with-parallel-storage>.
- [19] D. Yuan, Y. Yang, X. Liu, and J. Chen, "A data placement strategy in scientific cloud workflows," *Future Generation Computer Systems*, vol. 26, no. 8, pp. 1200–1214, 2010.
- [20] N. Maheshwari, R. Nanduri, and V. Varma, "Dynamic energy efficient data placement and cluster reconfiguration algorithm for MapReduce framework," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 119–127, 2012.
- [21] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [22] R. Chaiken, B. Jenkins, P. A. Larson, and B. Ramsey, "SCOPE: easy and efficient parallel processing of massive data sets," in *Proceedings of the VLDB Endowment VLDB Endowment Homepage Archive*, vol. 1-2, pp. 1265–1276, 2008.
- [23] D. Karger, E. Lehman, T. Leighton, and R. Panigrahy, "Consistent hashing and random trees: distributed caching protocols for relieving hot spots on the World Wide Web," in *Proceedings of the 29th*

- annual ACM Symposium on Theory of Computing (STOC'97)*, pp. 654–663, ACM, New York, NY, USA, 1999.
- [24] J. Dhok, N. Maheshwari, and V. Varma, “Learning based opportunistic admission control algorithm for MapReduce as a service,” in *Proceedings of the 3rd India Software Engineering Conference (ISEC'10)*, pp. 153–160, ACM, Mysore, India, February 2010.
 - [25] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, “Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility,” *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599–616, 2009.
 - [26] K. H. Kim, R. Buyya, and J. Kim, “Power aware scheduling of bag-of-tasks applications with deadline constraints on DVS-enabled clusters,” in *Proceedings of the 7th IEEE International Symposium on Cluster Computing and the Grid (CCGrid'07)*, pp. 541–548, Rio De Janeiro, Brazil, May 2007.

Research Article

Comprehensive Evaluation of Driver's Propensity Based on Evidence Theory

Jinglei Zhang,¹ Xiaoyuan Wang,^{1,2} Song Gao,¹ and Yiqing Song¹

¹ *School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255091, China*

² *Department of Civil and Environmental Engineering, School of Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

Correspondence should be addressed to Xiaoyuan Wang, wangxiaoyuan@sdut.edu.cn

Received 17 August 2012; Revised 29 November 2012; Accepted 2 December 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Jinglei Zhang 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 safety is related closely with driver's physiological and psychological characteristics. And the influence on traffic safety is represented as driver's propensity. Evidence theory is introduced to the evaluation system of driver's propensity in this paper, and it is utilized to combine the expert opinions, which can eliminate unavoidable uncertain elements in the traditional appraisal methods. The appraisal problems of subjective index can also be resolved by this method in the appraisal system. Results show that the method is objective and reasonable, and driver's propensity can be evaluated effectively.

1. Introduction

Transportation is the fundamental industry, which plays an important role in the development of the national economy. However, with the progress of urbanization and the popularity of vehicle, the transportation problem is seriously increasing; road traffic accidents have decreased in China in recent years, but the trend of traffic environment deterioration has not been resolved fundamentally. The internal factor of the driver himself is one of the main factors causing traffic accidents [1–5]. According to the research of international and China domestic scholars, driver's physiological-psychological characteristics are related closely with traffic safety, and the influence of psychological characteristics on driver's behavior is more important than the physiological characteristics [6–10]. The differences of driving behaviors are caused by differences of driver's gender, age, driving age, driving experience, and personality. All the differences can be summarized as driver's propensity difference.

Therefore, the impact of driver's physiological-psychological characteristics on traffic safety performs primarily as driver's propensity [11].

Driver's propensity is driver's attitude towards the objective reality traffic situations under the influence of various factors. It shows the psychological characteristics of the corresponding decision-making tendencies and reflects driver's psychological emotional state in the process of vehicle operation and movement. At present, China domestic and international experts focus on driver's physical and mental comprehensive evaluation [12]. A variety of evaluation methods are put forward, such as multiple regression, discriminant function, and neural network evaluation system. Wang and Zhang [13–16] had researched preliminarily driver's tendency on special traffic scenes, such as free flow and car following; Feng and Fang had researched cluster analysis of drivers characteristics evaluation [10]; Chen et al. had researched subjective judgment of driving tenseness and control of vehicle motion [11]; Wang et al. had researched reliability and safety analysis methodology for identification of drivers' erroneous actions [12]; Cai and Lin had researched modeling of operators' emotion and task performance in a virtual driving environment [17]. However, it is very difficult to evaluate driver's propensity because of driver's psychological changes and individual differences. Therefore, the research difficulty is to find a more reasonable method of driver's propensity evaluation.

Evidence theory is proposed firstly by Dempster in 1967 and is further developed by Shafer in 1976 [18]. It is also known as the Dempster/Shafer evidence theory (D-S evidence theory), which belongs to artificial intelligence areas. Evidence theory is used in expert systems in early days and can deal with uncertain information. In evidence theory, the evidence is not the actual evidence, but is part of the person's experience and knowledge and the results of people's observation and research. It not only emphasizes the objectivity of the evidence, but also emphasizes the subjectivity of the evidence [19]. The core of DS evidence theory is evidence combination rules. It can deal with the synthesis problem of vague and uncertain evidence. The theory can be applied to multiple attribute group decision-making evaluation method with "evidence information." The driver's propensity with a comprehensive evaluation of various test indexes can be evaluated and the accuracy of driver's propensity diagnosis can be improved through evidence theory.

2. Evidence Theory

Evidence theory is based on the merger of the evidence and the update of belief function. Its uncertainty is described by the identify framework, basic probability assignment function, trust function, the likelihood function, correlation, and so forth.

2.1. Identify Framework and Basic Trust Distribution Function

Based on the probability, the event in probability theory is extended to proposition, the event sets are extended to the proposition sets, and then the corresponding relationship between the proposition and sets is established by the evidence theory. The uncertainty problem of the proposition is transformed into the uncertainty problem of the set by introducing the trust function. Assume that there is a problem that needs to be judged, the complete set of all possible answers to this problem is expressed by Θ . So, any concerned proposition corresponds to a subset of the Θ , Θ is the identify framework, and the selection of Θ depends on the level of knowledge and understanding.

Definition 2.1. Make Θ as a recognition framework, the basic trust distribution function m is a collection from subset 2^Θ to $[0, 1]$, and A represents any subset of recognition framework Θ , $A \subseteq \Theta$, that meets

$$\begin{aligned} m(\emptyset) &= 0, \\ \sum_{A \subseteq \Theta} m(A) &= 1, \end{aligned} \quad (2.1)$$

where $m(A)$ is called basic trust distribution function of event A , and it means that the trust level is the evidence for the A .

2.2. Combination Rule of Evidence

D-S combination rule responses to evidence combined effects, which combines the independent evidence information from different sources to produce more reliable evidence information. Assuming that E_1 and E_2 are two pieces of evidence at the identify framework, the corresponding basic trust distribution functions are m_1 and m_2 , and focal element are A_1, \dots, A_k and B_1, \dots, B_l , respectively, if

$$K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j) < 1. \quad (2.2)$$

Then, these two groups of evidence can be combined and the combined basic probability assignment function $m: 2^\Theta \rightarrow [0, 1]$ meets

$$m(A) = \begin{cases} (1 - K)^{-1} \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j) & A \neq \emptyset, \\ 0 & A = \emptyset. \end{cases} \quad (2.3)$$

This is the famous D-S combination formula, where

$$K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j). \quad (2.4)$$

It reflects the conflict coefficient among various evidences. The trust function given by m is called the orthogonal summation of m_1 and m_2 , denoted by $m_1 \oplus m_2$.

3. Model Establishment

3.1. Multi-Index Evaluation Hierarchy Model

In order to improve the scientific nature of evaluation, the abstract goals are separated specifically into multilayered subobjectives. These sub objectives can be collectively referred to as indexes. Therefore, the object's evaluation can form a multilevel hierarchical structure of multi-index evaluation, the various evaluation indexes are synthesized from bottom to

top, and the value of the evaluation objects is judged. Finally, the final evaluation results can be gotten. In the evaluation process, the corresponding weight needs to be established for each index, and the weight can be determined by expert judgment, which reflects the quantitative distribution of the relative importance of each index. Make relative weights of E_i as w_i ($i = 1, \dots, r$) and meet

$$\sum_{i=1}^r w_i = 1 \quad (w_1, \dots, w_r \geq 0). \quad (3.1)$$

3.2. Discount Rate and Support Functions

The evidence theory usually uses the assumption of the identify framework Θ to undertake the uncertainty and also uses the support function to illustrate the uncertainty. But support functions may not reflect some special uncertainty of whole evidence, the uncertainty is α ($0 \leq \alpha \leq 1$) for whole evidence, and the parameter α reflects the discount rate given by the decision makers for the indexes evaluation results, so the evaluation results are not fully convinced. Therefore, the evaluation results of key index can be made as a benchmark, if $w_k = \max\{w_1, \dots, w_r\}$, then it is known as the key index.

Make $\beta_{kj}(1, \dots, n)$ as the probability of the key index E_k at state Θ_j decision makers determined the extent to believe that the state Θ_j will occur under this value β_{kj} . Given the corresponding discount rate, that is, $m_{kj} = \alpha\beta_{kj}$ ($j = 1, \dots, n$), m_{kj} , is used to express the support degree or the trust degree produced by the key indexes of decision makers. For unkey indexes E_i , the probability is β_{ij} at the state Θ_j . Analogously, the support degree can be produced by the discount of β_{ij} . Owing to the importance degree of the relative key index, E_k relative to E_i is w_i/w_k , so the $(w_i/w_k)\alpha$ is regarded as discount rate, and the $m_{ij} = (w_i/w_k)\alpha\beta_{ij}$ ($j = 1, \dots, n$) is used to express the thesis support degree of the unkey index. Aimed at the multi-index comprehensive evaluation in the model of level results, a corresponding evidential reasoning model can be established, and the corresponding evidence synthesis method can get the comprehensive basic probability function.

3.3. Driver's Propensity Evaluation System

Combining the driver physiological-psychological indexes in the literature [19] and the relevant indicators of driver's propensity psychological questionnaire in the literature [20], driver's propensity evaluation system can be constructed (e.g., Table 1) and driver's propensity can be measured. The first layer is the target layer, the second layer is the property secondary index, and the third layer is the factor layer index.

3.4. D-S Evidence Reasoning Model and Solving Method

First, set evaluation level of each factor level index as extraversion propensity, intermediacy propensity, and introversion propensity. Then, establish the identify framework according to the evidence; experts of related areas give the uncertainty subjective judgment of the identify elements. Based on different experiences and observations, different experts get different evidences necessarily in the identify framework Θ and obtain the basic probability assignment function. Each basic probability assignment function is synthesized

Table 1: Measurement index and precision.

Goals	Properties	Factors
Driving propensity	Reaction judgment indexes	Reaction time
		Speed estimated
		Complex reaction judgment
	Driving control indexes	Accelerator pedal intensity
		Brake pedal braking force
		Steering wheel grip force
Driving record indexes	Refueling frequency	
	Lane change frequency	
	Braking frequency	

by the combination rule of evidence to the compound, and the results given are an evaluation state of driver's propensity.

4. Instance Analysis

According to the driver's propensity index evaluation system mentioned, make $E = \{E_1, E_2, E_3\}$ express attribute level, $E_1 = \{e_{11}, e_{12}, e_{13}\}$, $E_2 = \{e_{21}, e_{22}, e_{23}\}$, $E_3 = \{e_{31}, e_{32}, e_{33}\}$, $E_1 \sim E_3$ express factor collection, $\Theta = \{\theta_1, \theta_2, \theta_3\}$ express comment collection of extraversion propensity, intermediacy propensity, and introversion propensity. Using expert judgment method to establish the relative weight of the layers of indexes, property layer weight is a collection $W = (w_1, w_2, w_3) = (0.3, 0.3, 0.4)$, weight vector of factors collection E_1, E_2, E_3 is $V_1 = (v_{11}, v_{12}, v_{13}) = (0.34, 0.33, 0.33)$, $V_2 = (v_{21}, v_{22}, v_{23}) = (0.35, 0.40, 0.25)$, $V_3 = (v_{31}, v_{32}, v_{33}) = (0.42, 0.36, 0.22)$.

10 experts use the extraversion propensity, intermediacy propensity, and introversion propensity to evaluate the indexes, and the evaluation results are shown as follows:

$$R_1 = \begin{bmatrix} 0.2 & 0.8 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.9 & 0 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 0.1 & 0.8 & 0.1 \\ 0.2 & 0.8 & 0 \\ 0.3 & 0.7 & 0 \end{bmatrix}, \quad R_3 = \begin{bmatrix} 0.2 & 0.7 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.8 & 0.2 \end{bmatrix}. \quad (4.1)$$

Among them, the matrix R_1 expresses that in the three factors of the property, "reaction judgment indexes," through the evaluation of "reaction time," the driver belongs to "extraversion propensity" by two experts, "intermediacy propensity" by eight experts, and "introversion propensity" by zero experts; through the evaluation of "speed estimate," the driver belongs to "extraversion propensity" by one expert, "intermediacy propensity" by eight experts, and "introversion propensity" by one expert; through the evaluation of "complex reaction judgment," the driver belongs to "extraversion propensity" by one expert, "intermediacy propensity" by nine experts, and "introversion propensity" by zero experts. The remaining two matrices have the same meaning. Now, take "reaction judgment indexes," for example, the evaluation of its driver's propensity is given in the calculation steps of evidential reasoning model. The evaluation information is shown in Table 2.

Transform the evaluation information of e_{11}, e_{12}, e_{13} into the basic probability assignment on the Θ . According to the relative weights between the factors, take $\alpha_1 = 0.8$ as

Table 2: Evaluation information of “reaction judgment indexes.”

Factor	Weights	Extraversion propensity	Intermediacy propensity	Introversion propensity
e_{11}	0.34	0.2	0.8	0
e_{12}	0.33	0.1	0.8	0.1
e_{13}	0.33	0.1	0.9	0

the discount rate of the key factor e_{11} , the discount rate of the nonkey factor is $(e_{12}/e_{11})\alpha_1 = 0.776$, and the discount rate of e_{13} is $(e_{13}/e_{11})\alpha_1 = 0.776$, then 3 basic probability assignments can be gotten. The synthetic results are shown in Table 3.

The synthesis results of E_2 and E_3 (the discount rate of the key factor was 0.8 and 0.9) are shown in Table 4.

Table 4 can be seen as an expert evaluation information matrix of the attribute set $E = \{E_1, E_2, E_3\}$, then transform it to basic probability assignment in the Θ and synthesize (the discount rate of the key factor is 1). The synthesis results are shown in Table 5.

The synthesis results show that, the support degree of extraversion propensity evidence is about 0.7% through evaluating object, intermediacy propensity is about 98.8%, and introversion propensity is about 0.4%. If the uncertainty is eliminated, driver’s propensity can be identified as intermediacy in theory.

5. Comparison of Evidence Synthesis Rules and Fuzzy Evaluation Method

For $E_1 = \{e_{11}, e_{12}, e_{13}\}$, its weight vector is $V_1 = (v_{11}, v_{12}, v_{13}) = (0.34, 0.33, 0.33)$, the comprehensive evaluation result of E_1 is $B_1 = V_1 R_1 = (0.134, 0.833, 0.033)$, and with the same reason, the comprehensive evaluation results of E_2 and E_3 are B_2 and B_3 , denoted as

$$R = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} 0.134 & 0.833 & 0.033 \\ 0.19 & 0.775 & 0.035 \\ 0.12 & 0.758 & 0.122 \end{bmatrix}, \quad (5.1)$$

then $B = WR = (0.1452, 0.7856, 0.0692)$, B is the comprehensive evaluation result. The driver’s propensity type is intermediacy according to the principle of maximum membership. It is the same as the result gotten by evidence theory reasoning mode. Because the fuzzy comprehensive evaluation is considered widely to be intuitive and reasonable, evidence theory reasoning model is verified to some extent.

6. Conclusion

Relationship between subgoals and overall goals is an inclusion relationship in fuzzy model, but in the evidence theory it is a “support” relationship, and subgoals are the support evidence. Obviously, evidence theory has a wider application range and more flexible application. In particular, the advantages of evidence theory can be better reflected when the individual property state is uncertain or unknown. In addition to the complete conflict, expert opinions cannot be synthesized by D-S theory, and other things can get better results.

Table 3: Synthesis conditions of “reaction judgment indexes” mass function.

Information source	Mass function	θ_1	θ_2	θ_3	Θ
e_{11}	m_{11}	0.16	0.64	0	0.2
e_{12}	m_{12}	0.0776	0.6208	0.0776	0.224
e_{13}	m_{13}	0.0776	0.6984	0	0.224
Synthesis	M_1	0.0333	0.9468	0.0051	0.0148

Table 4: Properties mass function.

Factors	Weights	Extraversion propensity	Intermediacy propensity	Introversion propensity	Θ
E_1	0.3	0.0333	0.9468	0.0051	0.0148
E_2	0.3	0.0835	0.8619	0.0103	0.0443
E_3	0.4	0.0539	0.8873	0.0392	0.0196

Table 5: Properties mass function synthesis situation.

Information source	Mass function	θ_1	θ_2	θ_3	Θ
E_1	M_1	0.025	0.7101	0.0038	0.2611
E_2	M_2	0.0626	0.6464	0.0077	0.2833
E_3	M_3	0.0539	0.8873	0.0392	0.0196
Synthesis		0.007	0.9875	0.0037	0.0018

Through establishing the identify framework, evidence theory is used to determine the basic probability assignment function; the qualitative “reaction judgment indexes, driving control indexes, and driving record indexes” are transformed into quantitative state. According to the different weights, the state of each index as well as the evaluation results of the experts is synthesized with combination rules, then the driver’s propensity by the synthesis function *mass* is determined. Aimed at the limitations of the experts, the evidence theory is used to study the uncertainty problem in this paper. The results show that the trust function constructed in this paper expresses uncertainty and obtains more accurate, reliable, and objective evaluation results.

Acknowledgments

This project is supported by National Natural Science Foundation of China (61074140), Natural Science Foundation of Shandong province (ZR2010FM007, ZR2011EEM034), Key Disciplines (Laboratory) Excellent Skeleton Teachers International Exchange Visitor Program of Shandong province, and Young Teacher Development Support Project of Shandong University of Technology.

References

- [1] W. H. Wang, W. Zhang, H. W. Guoi, H. Bubb, and K. Ikeuchi, “A safety-based behavioural approaching model with various driving characteristics,” *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [2] H. Bar-Gera and D. Shinar, “The tendency of drivers to pass other vehicles,” *Transportation Research F*, vol. 8, no. 6, pp. 429–439, 2005.

- [3] C. Pêcher, C. Lemercier, and J. M. Cellier, "Emotions drive attention: effects on driver's behaviour," *Safety Science*, vol. 47, no. 9, pp. 1254–1259, 2009.
- [4] O. Taubman, B. Ari, and E. Shay, "The association between risky driver and pedestrian behaviors: the case of Ultra-Orthodox Jewish road users," *Transportation Research F*, vol. 15, no. 2, pp. 188–195, 2012.
- [5] I. A. Kaysi and A. S. Abbany, "Modeling aggressive driver behavior at unsignalized intersections," *Accident Analysis and Prevention*, vol. 39, no. 4, pp. 671–678, 2007.
- [6] H. W. Warner, T. Özkan, T. Lajunen, and G. Tzamalouka, "Cross-cultural comparison of drivers' tendency to commit different aberrant driving behaviours," *Transportation Research F*, vol. 14, no. 5, pp. 390–399, 2011.
- [7] X. J. Wen, Y. D. Yang, and J. Fang, "Study on expressway operation speed," *Journal of Highway and Transportation Research and Development*, vol. 19, no. 1, pp. 81–83, 2000.
- [8] F. H. Yu, B. G. Wang, and X. J. Wu, *High Way Operation and Management*, People's Communications Publishing House, Beijing, China, 2000.
- [9] G. V. Barrett and C. L. Thornton, "Relationship between perceptual style and driver reaction to an emergency situation," *Journal of Applied Psychology*, vol. 52, no. 2, pp. 169–176, 1968.
- [10] Y. Q. Feng and C. Q. Fang, "Cluster analysis of drivers characteristics evaluation," *Communications Science and Technology Heilongjiang*, no. 11, pp. 161–163, 2007.
- [11] X. M. Chen, L. Gao, and S. B. Wu, "Research on subjective judgment of driving tenseness and control of vehicle motion," *Journal of Highway and Transportation Research and Development*, vol. 24, no. 8, pp. 144–148, 2007.
- [12] 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.
- [13] X. Y. Wang and J. L. Zhang, "Extraction and recognition methods of vehicle driving tendency feature based on driver-vehicle-environment dynamic data under car following," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1269–1281, 2011.
- [14] J. L. Zhang and X. Y. Wang, "The feature extraction and dynamic deduction method of vehicle driving tendency under time variable free flow condition," *Journal of Beijing Institute of Technology*, vol. 20, no. s1, pp. 127–133, 2011.
- [15] Y. Y. Zhang, X. Y. Wang, and J. L. Zhang, "Verification method of vehicle driving tendency recognition model under free flow," *Journal of Computer Application*, vol. 32, no. 2, pp. 578–580, 2012.
- [16] Y. Y. Zhang, X. Y. Wang, and J. L. Zhang, "The verification method of vehicle driving tendency recognition model based on driving simulated experiments under free flow," *Journal of Wuhan University of Technology*, vol. 33, no. 9, pp. 82–86, 2011.
- [17] H. Cai and Y. Lin, "Modeling of operators emotion and task performance in a virtual driving environment," *International Journal of Human Computer Studies*, vol. 69, no. 9, pp. 571–586, 2011.
- [18] F. B. Yang and X. X. Wang, *Combination Method of Conflict Evidence on Theory Evidence*, Defense Industry's Publishing House, Beijing, China, 2010.
- [19] H. Q. Jin, *Driving Adaptive*, Anhui People's Publishing House, Anhui, China, 1995.
- [20] Y. Y. Zhang, *Study on identification model and calculation method of driver's propensity based on dynamic driver-vehicle-environment data [M.S. thesis]*, Shandong University of Technology, Shandong, China, 2011.

Research Article

Concept Layout Model of Transportation Terminals

Li-ya Yao,¹ Li-shan Sun,² Wu-hong Wang,¹ and Hui Xiong¹

¹ Department of Transportation Engineering, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China

² Key Laboratory of Traffic Engineering, Beijing University of Technology, 100 Pingleyuan, Chaoyang District, Beijing 100124, China

Correspondence should be addressed to Li-ya Yao, yaoliya@bit.edu.cn

Received 28 April 2012; Revised 6 December 2012; Accepted 10 December 2012

Academic Editor: Klaus Bengler

Copyright © 2012 Li-ya 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.

Transportation terminal is the key node in transport systems. Efficient terminals can improve operation of passenger transportation networks, adjust the layout of public transportation networks, provide a passenger guidance system, and regulate the development of commercial forms, as well as optimize the assembly and distribution of modern logistic modes, among others. This study aims to clarify the relationship between the function and the structure of transportation terminals and establish the function layout design. The mapping mechanism of demand, function, and structure was analyzed, and a quantitative relationship between function and structure was obtained from a design perspective. Passenger demand and terminal structure were decomposed into several demand units and structural elements following the principle of reverse engineering. The relationship maps between these two kinds of elements were then analyzed. Function-oriented concept layout model of transportation terminals was established using the previous method. Thus, a technique in planning and design of transportation structures was proposed. Meaningful results were obtained from the optimization of transportation terminal facilities, which guide the design of the functional layout of transportation terminals and improve the development of urban passenger transportation systems.

1. Introduction

With the accelerated urbanization and motorization, construction of transportation terminals in major cities in China is gradually approaching the ideal. The operating efficiency of terminals, which are key nodes in the transportation network, directly influences the efficiency of transportation networks. However, the occurrence of passenger interleaving and long-distance walking due to deficiencies in the layout of transportation terminals result in a low operating efficiency. This problem is aggravated with the increase in transportation demand.

In many advanced cities, planning, design, and management of transportation terminals adaptive to their cities have been investigated since the 1950s to alleviate traffic congestion [1–3].

Batarliene and Jarašuniene [4] studied the interaction between different transport modes in transport terminals. Piccioni et al. [5] gave an application for facility location and optimal location models. Some scholars succeeded in the study of traffic characteristics of pedestrian. Lam et al. [6, 7] and Young Seth [8] obtained pedestrian walking speed at different facilities. Cheung and Lam [9, 10], Lee and Lam [11] and Delft [12] explained the pedestrian flow characteristics and route selection rule at subway stations and simulated facility service level. Progress in researches about characteristics and evolution law of the weaving behaviors of pedestrian flow in transport terminals has also been achieved. Henderson [13] analyzed the statistical characteristics of high density pedestrian flow. Satish et al. [14], Laxman et al. [15], and Lam et al. [16] studied the characteristics of pedestrian flow at certain transportation facilities. The relationship between pedestrian speed and density was studied by Ando et al. [17], Thompson and Marchant [18], Hughes [19], Hankin and Wright [20], and so on. All the achieved results formed a base for the planning, design, and management of transportation terminals. The study of public passenger transportation planning and design in China began in the 1990s. Among the successful efforts were the development of a technique that allows cooperation of public traffic and the subway, evaluation of joining coordination degree, streamline analysis of transfers, optimization of cohesive systems of transportation terminals, calculation of the main function of key facilities, and the optimized layout design of transportation terminals [21–23].

A number of local and international achievements in planning and design of transportation terminals have been reported; however, a traditional architectural design is usually used, without consideration and analysis of the traffic function of terminals as well as the matching facilities. Traditional architectural design cannot meet the demand of modern and efficient transportation terminals. The layout or design should be suitable for transportation structures such as transportation terminals. This study uses the decomposition and reconstitution mechanism in industry design to study the relationship maps between demand, function, and structure. The study also searches for a transportation terminal design based on demand, which will eventually provide a new method for the layout of transportation terminals.

The remainder of the paper is structured as follows. Section 2 briefly introduces the basic methods of concept layout model. Section 3 forms a concept layout model of transportation terminal based on mapping mechanisms among demand, function, and structure, followed by the conclusions in Section 4.

2. Basic Method

Apart from ordinary architectural structures and basic structural functions, transportation terminals play specific functions for traffic, business, and civil aviation. Thus, creating a layout of transportation terminals is a complex task. Such layout differs from those of ordinary architectural structures. This study proposes a new layout that satisfies the demand of both passengers and structures. The layout is developed from the perspective of passenger demand for efficient transportation terminals and maximum operating efficiency.

This study analyzes the different characteristics of passengers' demand and the structure of facilities. To clarify the uncertainty and multiplicity of the relationship maps between

demand and structure, the function layer was formed as the medium. A concept layout model of transportation terminals based on the relationship maps between demand, function, and structure was examined.

3. Concept Layout Model of Transportation Terminals

3.1. Mapping Mechanisms among Demand, Function, and Structure

Passengers engage in a series of activities inside and outside transportation terminals. These activities include purchasing tickets, security inspection, ticket checking, waiting, line transferring, boarding and alighting, and shopping. Each activity corresponds to an area within the terminal. However, the demand of passengers and the structure facilities of terminals are complicated. A passenger needs a series of structural facilities at certain times, and each structure can satisfy various passenger demands. To solve such problems, a functional transport layer has been established between passenger demands and structural facilities. This layer simulates human thinking.

Aimed at unifying demand, function, and structure, this paper first analyzed the original demand of passengers, the objective law of evolution of terminal function, and the characters of terminal structure facilities. For a certain passenger demand R_i , the corresponding function element F_i can be inversely decomposed. Facility elements of terminal structure can simultaneously be decomposed to form the mapping combination $C_i = \{R_i, F_i, S_i\}$ in Figure 1.

3.2. Concept Layout Model

Passenger demand in transportation terminals can be divided into several subdemands such as transportation, business, architecture, and civil aviation, among others, as shown in the first layer in Figure 2. Transportation demand is the core among the four sub-demands. Each subdemand can be divided into several demand units. For instance, let subdemand 1 represent transportation demand. Transportation demand can be divided into demand units such as transferring, waiting, buying tickets, parking, and coming in and out of the station.

To determine the functions of the demand units, corresponding structure facilities are needed. For example, to realize transfer demand, facilities such as transfer halls, transfer channels, railings, transfer stairs, autoescalators, automoving walkways, and elevators are necessary.

Thus, all facilities needed to realize the functions of the terminal are listed and then grouped into different substructures.

In Figure 2, the facilities are divided into four sub-structures, which consist of the following: inside and outside the area, fare collection system, transferring system, and platform area.

Substructure 1 includes the security inspection equipment, elevator, stairs, automatic moving walkway, automatic escalator, and channel.

Substructure 2 includes the wicket, automatic ticket-selling machine, autoinquiry machine, autorecharge system, pulling in and out of the station point, supplementary ticket desk, inquiry office, and railing.

Substructure 3 includes the transfer hall, transfer channel, railing, transfer stairs, auto-escalator, moving walkway, and elevator.

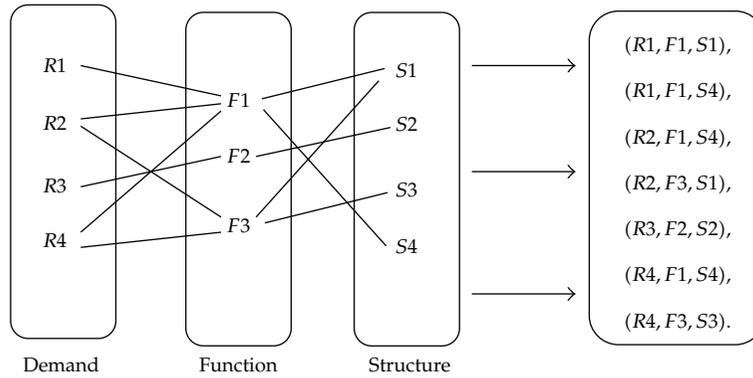


Figure 1: Relationships among demand, function, and structure.

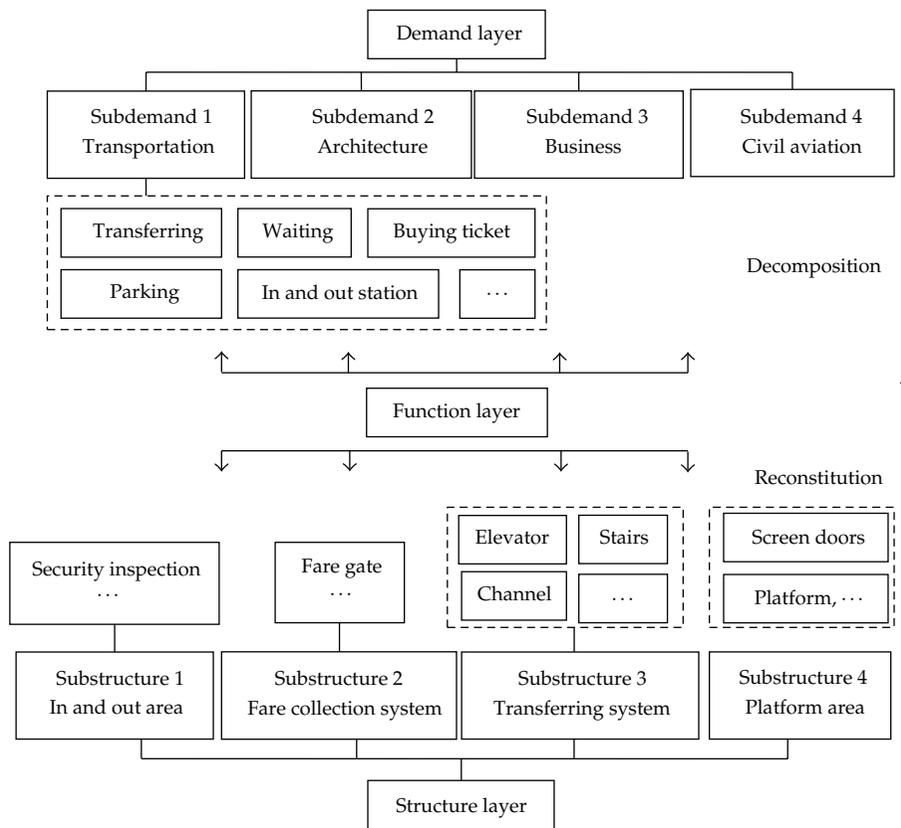


Figure 2: Decomposition and reconstitution of transportation concept layout model.

Substructure 4 includes the channel, elevator, auto-escalator, stairs, platform, and platform screen door.

Figure 2 shows the decomposition and reconstitution of transportation concept layout model.

3.3. Constraints of the Transportation Concept Layout Model

To form the detailed layout design, the growing elements must be calculated, including, the form, scale, relative position, and combination of cohesive primitive constraints of the facility. The key objective is to determine the constraints of all growing elements. With the aim of optimizing efficiency, this paper examines the calculation of the constraints of time utility, distance utility, and structure utility.

3.3.1. Constraint of Time Utility U_1

Time utility U_1 is the most important constraint in the calculation of growing elements of transportation terminals. U_1 is determined by the service level of facilities, passenger volume, adaptability relationship of the joining facilities, and so on. Time utility U_1 can be calculated by (3.1). The equation is explained by the arrival and departure of passengers in the terminal. Passengers are divided into four types: those arriving in mass and departing in mass, arriving individually and departing in mass, arriving in mass and departing individually, and arriving individually and departing individually

$$U_1 = \sum_{n=1}^4 Q_n (\bar{t}_{n1} + \bar{t}_{n2} + \bar{t}_{n3} + \bar{t}_{n4} + \bar{t}_{n5} + \bar{t}_{n6}), \quad (3.1)$$

where Q_n is the passenger volume of kind n , $n = 1, 2, 3, 4$; \bar{t}_{n1} is the average time required for type n passenger to buy tickets; \bar{t}_{n2} is the average time of security inspection for type n passenger; \bar{t}_{n3} is the average time required for type n passenger to arrive at the station; \bar{t}_{n4} is the average waiting time for type n passenger; \bar{t}_{n5} is the average time required for passenger n to depart from the station; and \bar{t}_{n6} is the average transfer time between different traffic modes or lines for type n passenger.

3.3.2. Constraint of Distance Utility U_2

Distance utility U_2 represents the basic constraint in the calculation of growing elements of transportation terminals. U_2 is determined by the relative position of facilities in passengers' walking streamline, as shown in (3.2)

$$U_2 = \sum_{m=1}^M Q_m \bar{d}_m, \quad (3.2)$$

where M is the number of transfer traffic modes; Q_m is the passenger volume of traffic mode m ; and \bar{d}_m is the average transfer walking distance of traffic mode m .

3.3.3. Constraint of Structure Utility U_3

Structure utility is the basic constraint in the calculation of growing elements of transportation terminals, which is determined by the facility plot ratio φ , organizational order of

streamline η and facility correlation degree θ , as shown in (3.3)

$$\begin{aligned} U_3 &= \varepsilon_1 + \eta\varepsilon_2 + \theta\varepsilon_3, \\ \varepsilon_1 + \varepsilon_2 + \varepsilon_3 &= 1, \end{aligned} \quad (3.3)$$

where $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\varepsilon_3 > 0$; φ is the facility plot ratio; η is the organizational order of streamline; θ is the degree of facility correlation; ε_1 is the weight of the facility plot ratio; ε_2 is the weight of organizational order of streamline; and ε_3 is the weight of the degree of facility correlation.

The facility plot ratio is determined by the average ratio of the effective facility utilization area S_I and the whole utilization area of the transportation terminal S_C , as shown in (3.4). The greater the plot ratio, the more reasonable the structure design

$$\varphi = \frac{1}{k} \sum_{i=1}^k \left(\frac{S_I}{S_C} \right)_i, \quad (3.4)$$

where k is the number of facility elements of terminals.

The organizational order of the streamline reflects the intereffect of each streamline, which is determined by the ratio of the total number of conflict points of the streamline H_0 and the total number of facility nodes of the streamline H_1 , as shown in

$$\eta = 1 - \frac{H_0}{H_1}. \quad (3.5)$$

Facility correlation degree is determined by the combined correlation value of each node of the streamline, as shown in (3.6). The greater the facility correlation degree, the more reasonable the structure design

$$\theta = \frac{1}{m} \sum_{i=1}^m \left[\frac{1}{r_i} \sum_{j=1}^{r_i} \left(\frac{n_{(j,j+1)} \cdot d_{(j,j+1)}}{\sum_{i=1}^m n_{(j,j+1)} \cdot d_{(j,j+1)}} \right) \right], \quad (3.6)$$

where m is the number of streamlines in the facility, r_i is the number of facilities at the node of the streamline i ; $n_{(j,j+1)}$ is the total number of passengers between the facility node j and the facility node $j + 1$; and $d_{(j,j+1)}$ is the walking distance from the facility node j and the facility node $j + 1$.

4. Conclusion

A transportation terminal design is established. The relationship between the function and the structure of transportation terminals is examined in this study. Following the principle of reverse engineering, the whole function and structure of the transportation terminal was decomposed into several demand units and elements of a facility structure. Transportation demand can be divided into subunit demands such as transferring, waiting, buying ticket, parking, and coming in and out of the station, among others. Facilities are divided into

four sub-structures: inside and outside the area, fare collection system, transfer system, and platform area. Furthermore, the calculation methods of the constraints of time, distance, and structure utilities are given. Based on this, the function-oriented concept layout model of transportation terminals is established to provide a new method for planning and designing transportation structures.

Future studies should focus on the quantitative description of the demand units and structure facilities.

Acknowledgments

This research was supported by the National Nature Science foundation of China (no. 51108028), Beijing Municipal Natural Science Foundation (no. 8122009), Programme of Introducing Talents of Discipline to Universities (no. B12022), and 973 Program (no. 2012CB725403).

References

- [1] T. Tomiyama, *General Design Theory and Its Extensions and Applications*, Universal Design Theory, 1998.
- [2] M. A. Rosenman, "An exploration into evolutionary models for non-routine design," *Artificial Intelligence in Engineering*, vol. 11, no. 3, pp. 287–293, 1997.
- [3] J. S. Gero and V. A. Kazakov, "Evolving design genes in space layout planning problems," *Artificial Intelligence in Engineering*, vol. 12, no. 3, pp. 163–176, 1998.
- [4] N. Batarliene and A. Jarašuniene, "Research on advanced technologies and their efficiency in the process of interactions between different transport modes in the terminal," *Transport*, vol. 24, no. 2, pp. 129–134, 2009.
- [5] C. Piccioni, F. Antomazzi, and A. Musso, "Locating combined road-railway transport terminals: an application for facility location and optimal location models," *Ingegneria Ferroviaria*, vol. 65, no. 7-8, pp. 625–649, 2010.
- [6] W. H. K. Lam and C. Y. Cheung, "Pedestrian speed/flow relationships for walking facilities in Hong Kong," *Journal of Transportation Engineering*, vol. 126, no. 4, pp. 343–349, 2000.
- [7] 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.
- [8] B. Young Seth, "Evaluation of pedestrian walking speeds in airport terminals," in *Proceedings of the 78th TRB Annual Meeting*, pp. 20–26, National Research Council, Washington, DC, USA, 1999.
- [9] C. Y. Cheung and W. H. K. Lam, "A study of the Bi-directional pedestrian flow characteristics in the Hong Kong mass transit railway stations," *Journal of Transportation Engineering*, vol. 2, no. 5, pp. 1607–1619, 1997.
- [10] 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.
- [11] J. Y. S. Lee and W. H. K. Lam, "Levels of service for stairway in Hong Kong underground stations," *Journal of Transportation Engineering*, vol. 129, no. 2, pp. 196–202, 2003.
- [12] W. Delft, *Modeling Passenger Flows in Public Transport Facilities*, Delft University, Delft, The Netherlands, 2004.
- [13] L. F. Henderson, "The statistics of crowd fluids," *Nature*, vol. 229, no. 5284, pp. 381–383, 1971.
- [14] U. Satish, H. Samiul, and A. H. M. Abdul, "Random parameter model used to explain effects of built-environment characteristics on pedestrian crash," *Transportation Research Record*, vol. 2011, no. 2237, pp. 98–106, 2011.
- [15] K. K. Laxman, R. Rastogi, and S. Chandra, "Pedestrian flow characteristics in mixed traffic conditions," *Journal of Urban Planning and Development*, vol. 136, no. 1, Article ID 004001QUP, pp. 23–33, 2010.
- [16] W. H. K. Lam, J. F. Morrall, and H. Ho, "Pedestrian flow characteristics in Hong Kong," *Transportation Research Record*, vol. 1995, no. 1487, pp. 56–62, 1995.

- [17] K. Ando, H. Ota, and T. Oki, "Forecasting the flow of people," *Railway Research Review*, vol. 45, no. 8, pp. 8–14, 1988.
- [18] P. A. Thompson and E. W. Marchant, "A computer model for the evacuation of large building populations," *Fire Safety Journal*, vol. 24, no. 2, pp. 131–148, 1995.
- [19] R. L. Hughes, "The flow of large crowds of pedestrians," *Mathematics and Computers in Simulation*, vol. 53, no. 4–6, pp. 367–370, 2000.
- [20] B. D. Hankin and R. A. Wright, "Passenger flow in subways," *Operational Research Quarterly*, vol. 9, no. 2, pp. 81–88, 1958.
- [21] X. Qing, "Developing an urban transportation system centered with public transportation-Thinking about the transportation situation of CBD in Guangzhou," *Journal of Guangzhou University*, vol. 3, no. 3, pp. 243–246, 2004.
- [22] Z. Shuai, *The Optimal Design Study of Urban Rail Transit Hub Internal Space Traffic Streamline*, Beijing Jiaotong University, 2011.
- [23] L. J. Yan, Z. Zuo, Z. Z. Yang, and H. G. Kang, "Optimization of location of transportation terminal with super-network," *Journal of Harbin Institute of Technology*, vol. 38, no. 8, pp. 1344–1346, 2006.

Research Article

Offset Optimization Based on Queue Length Constraint for Saturated Arterial Intersections

Xianmin Song,¹ Pengfei Tao,¹ Ligang Chen,¹ and Dianhai Wang²

¹ Transportation College, Jilin University, Changchun 130022, China

² Civil Engineering and Architecture College, Zhejiang University, Hangzhou 310058, China

Correspondence should be addressed to Xianmin Song, songxm@jlu.edu.cn

Received 17 August 2012; Revised 27 November 2012; Accepted 4 December 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Xianmin Song 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.

Offset optimization is of critical importance to the traffic control system, especially when spillovers appear. In order to avoid vehicle queue spillovers, an arterial offset optimization model was presented in saturated arterial intersections based on minimizing the queue length over the whole duration of the saturated traffic environment. The paper uses the shockwave theory to analyze the queue evolution process of the intersection approach under the saturated traffic environment. Then through establishing and analyzing a function relationship between offset and the maximum queue length per cycle, a mapping model of offset and maximum queue length was established in the saturated condition. The validity and sensitivity of this model were tested by the VISSIM simulation environment. Finally, results showed that when volumes ratios are 0.525–0.6, adjusting offset reasonably under the saturated condition could decrease the queue length and effectively improve the vehicle operating efficiency.

1. Introduction

In arterials systems, reasonable adoption of the coordinated control could decrease vehicle delays and stops. Offset is an important parameter of the arterial signal coordinate control, which decides the effect of the coordinated control for adjacent intersections. Now, there are mainly two kinds of the offset optimization [1–3]: (1) Maximum Green Wave Band, setting offset in order to obtain green wave travel time (maximize ratio of the green wave band width to cycle length). It includes Graphical Method, Mathematical Analysis Method, and Max Band Method; (2) minimize vehicle delays and/or stops, optimizing offset by an index function relation between the vehicle delays (stops) and offset.

The above arterial coordination methods play an important role in the practical traffic management. However, these methods are mainly designed to the undersaturated coordination, namely, link supply can satisfy upstream input demand [4]. With the increasing of

vehicle year by year, the saturated state of the arterial system becomes more common. If we use the above methods to analyze the saturated condition, we would obtain misleading results. Therefore, the study of offset optimization at the saturated arterial has important theoretical significance and application value.

When traffic demand exceeds road capacity or intersection capacity, the saturated condition appears, and vehicle queues continue to increase. These queues can overflow the storage capacity of the road and physically block intersections and gradually spread to the surrounding intersections. Implementing signal control policies designed for the undersaturated condition is ineffective and even counters product to the saturated condition [5–7]. Therefore, carrying on traffic dispersion and decreasing the queue length are major signal control objectives of the saturated environment. Now, the study of the saturated arterial signal coordinated control mainly included the following three categories: (1) maximum system capacity, based on an optimization objective of maximum release amount of system; Khatib and Judd [8] gave a corresponding control strategy, and Ghassan and Rahim [9] conducted coordinate control based on a control objective of the system turn over volumes. (2) Minimize vehicle delays or stops: an optimization strategy [10–12] mainly analyzes traffic delay of traveling team over the whole duration for the saturated traffic environment, then an arterial offset optimization model was presented based on an objective of minimizing delay or queue length. (3) Control queue length: Chang and Lin [13] proposed a saturated traffic control method, which managed the queuing dissipation and formation in subarea sections, and Lieberman et al. [14] proposed a method to optimize queue length based on the saturated arterial coordinate control strategy. Wang et al. [15, 16] put pedestrian safety as an important factor in coordinated system. However, this paper is focused on the theoretical models per cycle; there is no optimal control of the maximum queue length for multiple consecutive cycles.

According to an objective of minimizing queue length over the whole duration of the saturated traffic environment, the paper presented a method to optimize offset of the saturated arterial through analyzing queue evolution process based on shockwave theory. Through establishing and analyzing functions relationship between offset and maximum queue length of single cycle, a mapping model of offset and maximum queue length is established in the saturated condition. Finally, validity and sensitivity of the model are tested in VISSIM software.

2. The Queue Evolution Process of the Saturated Traffic Environment

2.1. State Description of the Saturated Arterial

When traffic demand is much more than the capacity at saturated arterial intersections, the number of in-flowing vehicles exceeds those discharging from the approach, and queue length will gradually increase at downstream intersection. Assuming the signal timing parameters remain fixed, this current state of the approach will sustainably develop, then the long queuing will spill over to the upstream intersection. Our study is based on the following basic assumptions.

(1) Saturated arterial intersections are typical four phases, and the phase sequence remains consistent and fixed. Every cycle is composed of the green time and red time without considering the yellow time. Right-turn flows at the adjacent intersection are less and without signal control, so that its effect can be neglected. Figure 1 shows the main direction of traffic flows and phase sequence at saturated arterial adjacent intersections, in which east-west through phase is coordination phase.

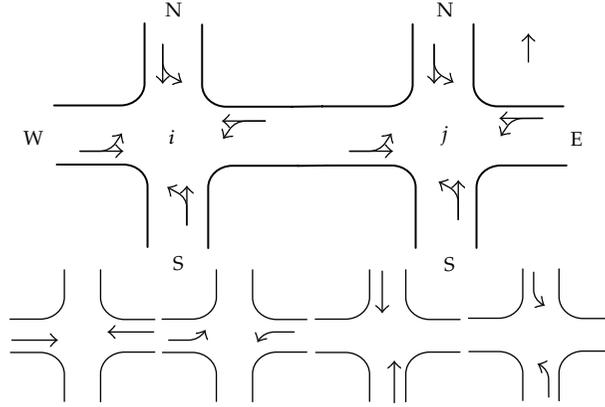


Figure 1: Main direction of traffic flows and phase-sequence at saturated arterial adjacent intersections.

As shown in Figure 1, i is upstream intersection, while j is downstream intersection. Arrival vehicle of intersection j is downstream that consists of two parts: west approach through flows and north approach left-turn flows of upstream intersection. The number of arrival vehicles per cycle of downstream intersection is

$$Q_W(j) = D_W(i)[1 - p_W(i)]g_{WES}(i) + D_N(i)p_N(i)g_{NEL}(i), \quad (2.1)$$

where $Q_W(j)$ is the number of arrival vehicles per cycle of downstream intersection; $D_W(i)$ and $D_N(i)$ are west approach and north approach discharge rate of intersection j ; $p_W(i)$ and $p_N(i)$ are west approach and north approach turning movement proportions of intersection j ; $g_{WES}(i)$ and $g_{NEL}(i)$ are east-west straight and south-north left-turn phase green time.

(2) Coordination phase green time of upstream intersection is adequate. When green time starts, the platoon firstly discharges with saturation flow rate, and later discharges with average arrival rate. If we ignore the discreteness of platoon to reach downstream intersection, the travel platoon will be pulse arriving to downstream:

$$D(i) = \begin{cases} S, & 0 < t \leq g_s(i), \\ q_W(i), & g_s(i) < t \leq g_{WES}(i), \end{cases} \quad (2.2)$$

where $q_W(i)$ and $D(i)$ are the arrival and discharge rates of upstream intersection coordination phase; S is the saturation flow rate; $g_s(i)$ is the duration of saturation discharging of upstream intersection coordination phase. The duration of saturation discharging is

$$g_s(i) = \frac{(C - g_{WES}(i)) \cdot q_W(i)}{S - q_W(i)}. \quad (2.3)$$

Then there are three femoral pulses of the travel platoons, as Figure 2 showed. Substituting (2.2) into (2.1), then

$$Q_W(j) = [Sg_s(i) + q_W(i)(g_{WES}(i) - g_s(i))][1 - p_W(i)]g_{WES}(i) + D_N(i)p_N(i)g_{NEL}(i). \quad (2.4)$$

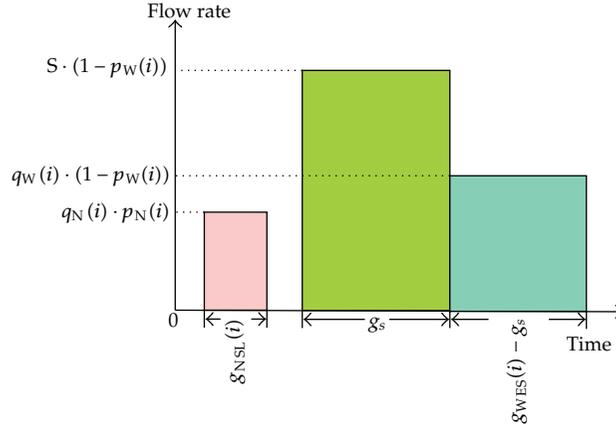


Figure 2: Pulse arrival flows schema of downstream intersection.

(3) Maximum cycle length model to determine intersection cycle [17]:

$$C = \sqrt{\frac{LT}{(1-\omega)Y'}} \quad (2.5)$$

where C is the cycle time; L is the total loss time; T is the duration of saturated conditions; Y is the total flow ratio.

(4) The common cycle length of the saturated arterial is determined by traditional methods [18], and the numbers of the saturated cycles can be calculated as follows:

$$\delta = \left\lceil \frac{T}{C_o} \right\rceil, \quad (2.6)$$

where C_o is the common cycle length.

(5) Green time of each intersection allocated by equal saturation principle.

2.2. Queue Evolution Analysis

At a signalized intersection, multiple shock waves are generated due to the stop-and-go traffic caused by signal changes. As indicated in Figure 3, when upstream intersection green time starts, the platoon firstly discharging of saturation flow rate S discharges to downstream intersection. The head vehicle of platoon forced to stop at downstream intersection because of red time, at which it creates different flows and density conditions between the arrival and stopped traffic, forms a queuing shockwave:

$$v_1 = \frac{q_a^1}{k_j - k_a^1}, \quad (2.7)$$

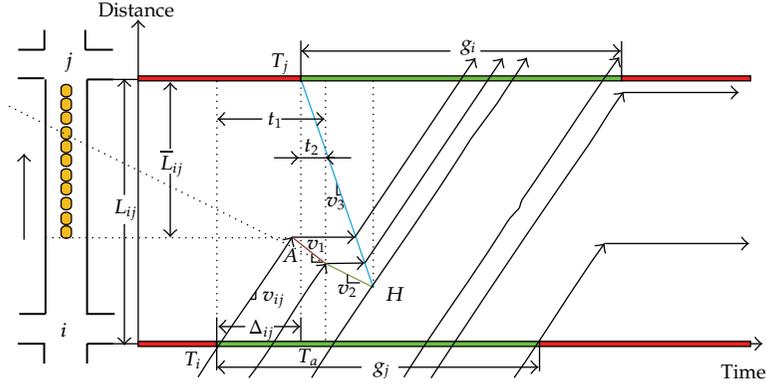


Figure 3: Traffic flow temporal-spatial graph at saturated arterial adjacent intersections.

where k_j represents the jammed density; q_a^1 and k_a^1 are the average arrival flow rates and density of first part flows; q_a^1 and k_a^1 can be calculated as [19]

$$q_a^1 = S \cdot \frac{g_s(i)}{C_o}, \quad (2.8)$$

$$k_a^1 = \frac{q_a^1}{v_{ij}}$$

Substituting (2.8) into (2.7), then

$$v_1 = \frac{S \cdot h \cdot g_s(i) \cdot v_{ij}}{1000C_o \cdot v_{ij} - S \cdot h \cdot g_s(i)}, \quad (2.9)$$

where h is mean queue discharge headway.

When arrival vehicles are of the saturation flow rate all joining the queue, the second part flows will continue to queue, forming a queuing shockwave:

$$v_2 = \frac{q_w(i) \cdot h \cdot (g_{WES}(i) - g_s(i)) \cdot v_{ij}}{1000C_o \cdot v_{ij} - q_w(i) \cdot h \cdot (g_{WES}(i) - g_s(i))}. \quad (2.10)$$

Similarly, at the beginning of the downstream intersection green, vehicles begin to discharge of the saturation flow rate forming a discharge shockwave:

$$v_3 = \frac{S \cdot h \cdot g_s(j) \cdot v_{ij}}{1000C_o \cdot v_{ij} - S \cdot h \cdot g_s(j)}. \quad (2.11)$$

v_1 , v_2 , v_3 all moved upstream of the intersection and finally intersected at point H , which is the point of maximum queue length.

3. Offset Optimization for the Saturated Arterial

3.1. Single Cycle Function Relationship between Offset and Maximum Queue Length

Figure 3 shows the operating condition of adjacent intersections flows. At the beginning of the downstream intersection green, head vehicle is of average speed v_{ij} traveling to the downstream intersection and joins the end of original queue in which speed reduces to 0. Until the discharge shockwave of downstream intersection spreads to this location, head vehicle continues to travel downstream of average speed v_{ij} . Subsequently the following head vehicles that arrive before point H are in such process ($v_{ij} \rightarrow 0 \rightarrow v_{ij}$) running in the link. After point H , arrive vehicles travel downstream of average speed unimpeded. The above process is shown in Figure 3. According to Figure 3, offset can be represented as

$$\phi_{ij} = t_1 - t_2, \quad (3.1)$$

where ϕ_{ij} is the offset of intersection i, j . t_1 and t_2 can be calculated as

$$\begin{aligned} t_1 &= g_s(i) + \frac{L_{ij} - (Q_{ij} + S \cdot g_s(i)) \cdot h}{v_{ij}}, \\ t_2 &= \frac{(Q_{ij} + S \cdot g_s(i)) \cdot h}{v_2} - N_{ij} \cdot h \left(\frac{1}{v_2} - \frac{1}{v_3} \right), \end{aligned} \quad (3.2)$$

where Q_{ij} is the original queue numbers of intersection i, j , veh; N_{ij} is the maximum queue numbers of intersection i, j , veh. Substituting (3.2) into (3.1), then

$$\phi_{ij} = g_s(i) + \frac{L_{ij}}{v_{ij}} + N_{ij} \cdot h \left(\frac{1}{v_2} - \frac{1}{v_3} \right) - (Q_{ij} + S \cdot g_s(i)) \cdot h \left(\frac{1}{v_{ij}} + \frac{1}{v_2} \right). \quad (3.3)$$

The maximum queue length of intersection i, j can be represented as (15) is a computation model of maximum queue numbers per cycle. This model should be operated with the following two situations.

(1) When $\phi_{ij} > (L_{ij} - Q_{ij} \cdot h) / v_{ij} - Q_{ij} \cdot h / v_3$, maximum queue numbers per cycle and offset, common cycle, green time, original queue numbers, arrival rate, and discharge rate have a close relationship.

(2) When $\phi_{ij} \leq (L_{ij} - Q_{ij} \cdot h) / v_{ij} - Q_{ij} \cdot h / v_3$, maximum queue numbers equal to original queue numbers, arrival rate. Although this time of queue numbers is the shortest, it will result in the loss of downstream intersection green, not considered this situation:

$$N_{ij} = \begin{cases} \frac{[\phi_{ij} - g_s(i) - L_{ij} / v_{ij} + (Q_{ij} + S \cdot g_s(i)) \cdot h (1 / v_{ij} + 1 / v_2)]}{h \cdot (1 / v_2 - 1 / v_3)}, & \phi_{ij} > \frac{L_{ij} - Q_{ij} \cdot h}{v_{ij}} - \frac{Q_{ij} \cdot h}{v_3}, \\ Q_{ij}, & \phi_{ij} \leq \frac{L_{ij} - Q_{ij} \cdot h}{v_{ij}} - \frac{Q_{ij} \cdot h}{v_3}. \end{cases} \quad (3.4)$$

3.2. Computation Model of Queue Length Diversification

Under the saturated condition, if the signal timing is fixed, the arrival rate and discharge rate stability, the maximum queue numbers, and original queue numbers are greater than previous cycles, and the relation of queue lengths between cycle k and $k + 1$ can be represented by the following equations:

$$\begin{aligned} Q_{ij}(k+1) &\geq Q_{ij}(k), \\ N_{ij}(k+1) &\geq N_{ij}(k). \end{aligned} \quad (3.5)$$

The maximum queue numbers and original queue numbers of intersection i, j are increased during the saturation period cycle by cycle, and reach the maximum in the last saturated cycle. We can get the law of queue length changes between adjacent cycles as shown in Figure 4. Ignore the disturbance of arrival and discharge; variable quantity of the original queue numbers can be calculated as

$$\begin{aligned} \Delta Q_{ij} = & [S \cdot g_s(i) + q_W(i) \cdot (g_{WES}(i) - g_s(i))] \cdot (1 - p_W(i)) + q_N(i) \cdot p_N(i) \cdot g_{NSL}(i) \\ & - [S \cdot g_s(j) + q_W(j) \cdot (g_{WES}(j) - g_s(j))] \cdot (1 - p_W(j)). \end{aligned} \quad (3.6)$$

An algorithm for solving the variable quantity of maximum queue numbers model is arranged in the following steps.

Step 1. By introducing parameter $t_s(k)$, which represents the residence time of head vehicle in the first part flows joining the end of original queue,

$$t_s(k) = \Delta_{ij}(k) + \frac{Q_{ij}(k) \cdot h}{v_1} - \frac{L_{ij} - Q_{ij}(k) \cdot h}{v_{ij}}. \quad (3.7)$$

There are continuous vehicles from upstream joining the queue; assuming queue residence time is $t_{s+1}(k), t_{s+2}(k), \dots, t_{s+n-1}(k)$, then they are negatively correlated:

$$t_{s+1}(k) > t_{s+2}(k) > \dots > t_{s+n-1}(k) > 0. \quad (3.8)$$

Step 2. The queue numbers of first part flows from upstream in cycle k are

$$n_1(k) = S \cdot g_s(i) \cdot (1 - p_W(i)). \quad (3.9)$$

Step 3. By introducing parameter $t'_s(k)$, which represents the residence time of head vehicle in the second part flows joining the end of the original queue,

$$t'_s(k) = t_s(k) + n_1(k) \cdot h \left(\frac{1}{v_3} - \frac{1}{v_1} \right). \quad (3.10)$$

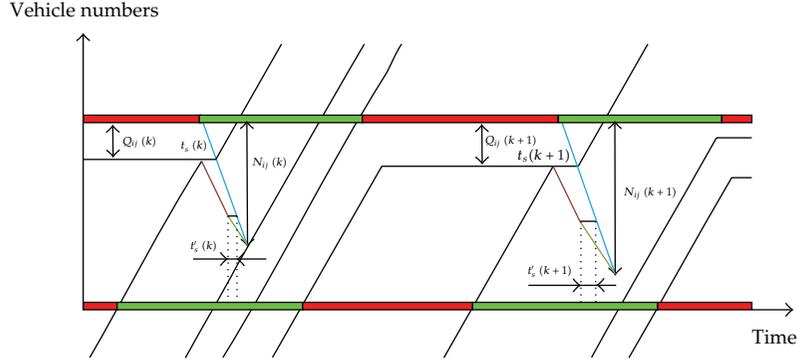


Figure 4: Arterial traffic flow temporal-spatial graph in multisaturated cycle.

Step 4. The queue numbers of total flows from upstream in cycle k are:

$$\begin{aligned} t'_s(k) &= (n(k) - n_1(k)) \cdot h \cdot \left(\frac{1}{v_2} - \frac{1}{v_3} \right) \\ \implies n(k) &= n_1(k) + \frac{t'_s(k) \cdot v_2 \cdot v_3}{(v_3 - v_2) \cdot h}. \end{aligned} \quad (3.11)$$

Then the maximum queue numbers in cycle k can be calculated as

$$N_{ij}(k) = Q_{ij}(k) + n_1(k) + \frac{t'_s(k) \cdot v_2 \cdot v_3}{(v_3 - v_2) \cdot h}. \quad (3.12)$$

Step 5. We can get the variable quantity of maximum queue numbers according to (3.12):

$$\begin{aligned} \Delta N_{ij} &= N_{ij}(k+1) - N_{ij}(k) \\ &= \Delta Q_{ij} + [n_1(k+1) - n_1(k)] + \frac{v_2 \cdot v_3}{(v_3 - v_2) \cdot h} [t'_s(k+1) - t'_s(k)], \end{aligned} \quad (3.13)$$

then

$$\Delta N_{ij} = \Delta Q_{ij} \cdot \left[1 + \frac{v_2 \cdot v_3}{v_3 - v_2} \left(\frac{1}{v_1} + \frac{1}{v_{ij}} \right) \right]. \quad (3.14)$$

The maximum queue numbers and original queue numbers of the last saturated cycle can be represented by the following equation:

$$\begin{aligned} Q_{ij}(\delta) &= Q_{ij}(1) + (\delta - 1)\Delta Q_{ij}, \\ N_{ij}(\delta) &= N_{ij}(1) + (\delta - 1)\Delta N_{ij}. \end{aligned} \quad (3.15)$$

3.3. Offset Optimization Model

Through building a single cycle function relationship between offset and maximum queue length, a computation model of queue length diversification each cycle during duration of saturated, we propose an arterial offset optimization model based on an objective of minimizing queue length over the whole duration of the saturated traffic environment, avoiding queue spillover.

Equation (3.16) builds the relationship between $N_{ij}(\omega)$ and ϕ_{ij} . If other parameters are known, an optimal offset ϕ_{ij}^{opt} by simple iterative method can be obtained:

$$\begin{aligned} \min \quad N_{ij}(\omega) &= \frac{[\phi_{ij} - g_s(i) - L_{ij}/v_{ij} + (Q_{ij} + S \cdot g_s(i)) \cdot h(1/v_{ij} + 1/v_2)]}{h \cdot (1/v_2 - 1/v_3)} \\ &\quad + (\delta - 1) \cdot \Delta Q_{ij} \cdot \left[1 + \frac{v_2 \cdot v_3}{v_3 - v_2} \left(\frac{1}{v_1} + \frac{1}{v_{ij}} \right) \right], \end{aligned} \quad (3.16)$$

$$\begin{aligned} \text{s.t.} \quad x_{\text{WES}} &= \frac{q_{\text{W}} C_o}{g_{\text{WES}} S} > 0.9 & (a) \\ N_{ij} &\geq Q_{ij} & (b) \\ -g_{\text{WES}}(j) &\leq \phi_{ij} \leq C - g_{\text{WES}} & (c) \end{aligned}$$

4. Simulation Testing and Analysis

In order to verify the application effect of the Saturated Arterial Optimization method, we simulate traffic operation state of two intersections adopting single point control (Scheme 1), coordination control (Scheme 2), and optimization coordination control (Scheme 3) by the VISSIM software, also comparative analysis of the output evaluation of these simulation scenarios.

4.1. Simulation Testing Scheme

Simulation Testing Arterial Schematic as shown in Figure 5: all approaches are with dedicated left turn traffic lane and they do not have the slip/slide right turn lane because we have not taken right-turning vehicles into consideration when modeling. Designing an arterial of approach $L_{ij} = 400$ m, mean queue discharge headway $h = 6$ m, mean travel speed $v_{ij} = 40$ km/h, saturation flow rate $S = 1800$ veh/h, all entry turning movement proportions of intersection $p_{\text{W}} = p_{\text{E}} = 0.15$, $p_{\text{N}} = p_{\text{S}} = 0.3$.

All entry volumes and approach lengths of arterial are showed in Table 1: assumed road section length between intersection i and j is 400 m, and assumed related traffic volumes are 1794 veh/h and 128 veh/h.

Assume the saturated duration of two intersections is 1200 s, total loss time $L = 12$ s, the timing scheme of three different simulation scenarios can be obtained (see Table 2).

4.2. Discussion of Simulation Results

Output the maximum queue length, and mean queue length, mean vehicle delays of two intersections with the three timing schemes in the last saturated cycle.

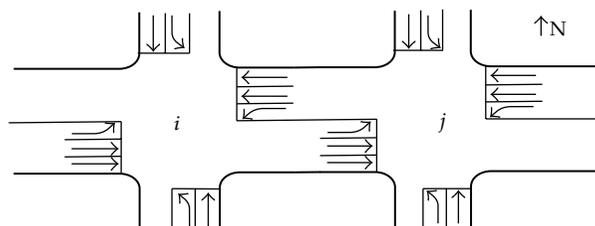


Figure 5: Simulation testing arterial schematic.

Table 1: All entry flows and approach length of the arterial.

Direction	Intersection <i>i</i>		Intersection <i>j</i>	
	Volumes (veh/h)	Approach length (m)	Volumes (veh/h)	Approach length (m)
East approach				
Through	—	325	1738	400
Left turn	—	—	242	—
West approach				
Through	1794	400	—	400
Left turn	269	—	—	—
North approach				
Through	298	275	400	425
Left turn	128	—	171	—
South approach				
Through	272	300	498	355
Left turn	119	—	213	—

As shown in Table 3, when using optimized offset, the maximum queue length of link is 263 m, less than the maximum queue length of Scheme 1 and Scheme 2, which decreased by 10.8% compared with Scheme 2. Mean queue length is also decreased by 15.3%. The maximum queue length simulation results of objective link during the saturated period are shown in Figure 6. In addition, when running the timing parameters of Scheme 3, mean vehicle delays of coordination phase remain fixed compared with the first two schemes. In particular, maximum queue length, mean queue length, and mean vehicle delays of noncoordination phase have different degrees of improvement. Therefore, to ensure the operation parameters of competitive phase not deteriorated, the offset which is calculated by our optimization method can effectively reduce the maximum queue length and prevents queuing spillover in the approach of coordination phase.

4.3. Sensitivity Analysis

There is a function of offset and maximum queue length, and determining the offset reasonably can reduce the maximum queue length. Whether this optimization of different arrival rates significant or not, which requires the sensitivity analysis.

As the shown in Figure 7, with volumes increasing, the effectiveness which adjusts offset to reduce maximum queue length is more significant. But not volumes the larger the better, when volumes ratios (the ratio of volume and saturation flow rate) are 0.525–0.6, avoiding queuing spillover by optimizing offset; volumes ratio above 0.6, adjusting offset

Table 2: Signal timing parameters of the 3 simulated schemes.

	Scheme 1		Scheme 2		Scheme 3	
	Intersection	Intersection	Intersection	Intersection	Intersection	Intersection
	<i>i</i>	<i>j</i>	<i>i</i>	<i>j</i>	<i>i</i>	<i>j</i>
Cycle length (m)	193	176	193	193	193	193
W-E through green (s)	102	91	102	93	102	93
W-E left-turn green (s)	31	27	31	28	31	28
N-S through green (s)	34	41	34	42	34	42
N-S left-turn green (s)	15	17	15	18	15	18
Offset (s)	—		36		-31	

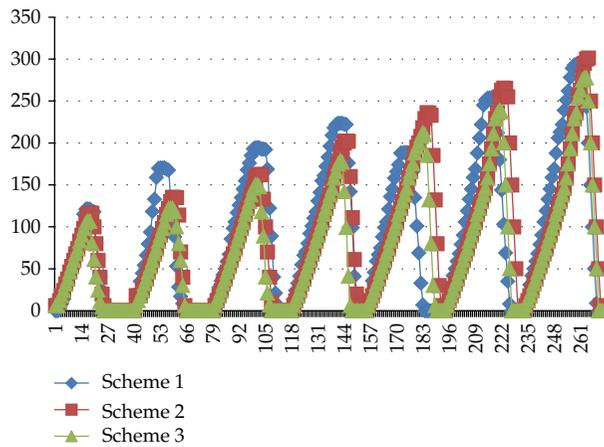


Figure 6: Simulation results of link maximum queue length.

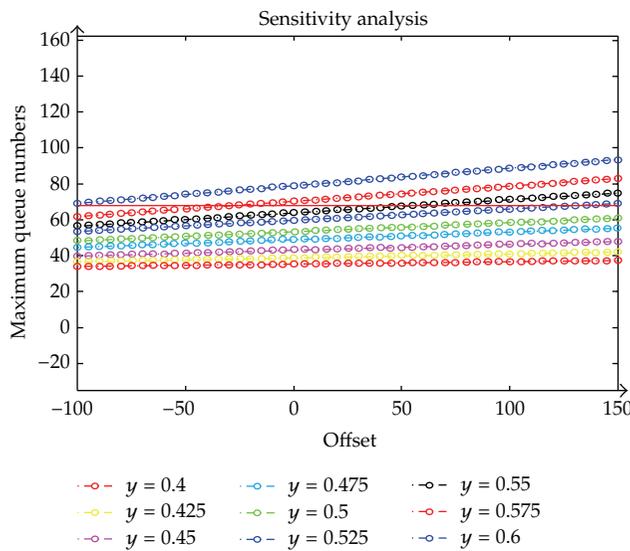


Figure 7: The relation between offset and maximum queue length.

Table 3: Outputs parameters of the simulations.

	Scheme 1	Scheme 2	Scheme 3
Intersection <i>i</i>			
Coordination phase			
Maximum queue length (m)	221	208	202
Mean queue length (m)	113	101	96
Mean vehicle delays (s)	108	100	88
Mean value of noncoordination phase			
Maximum queue length (m)	121	98	82
Mean queue length (m)	64	42	35
Mean vehicle delays (s)	78	64	58
Intersection <i>j</i>			
Coordination phase			
Maximum queue length (m)	289	295	263
Mean queue length (m)	162	170	134
Mean vehicle delays (s)	125	127	128
Mean value of noncoordination phase			
Maximum queue length (m)	141	132	122
Mean queue length (m)	89	75	66
Mean vehicle delays (s)	88	83	80

alone cannot reduce the queue length, spillover is inevitable, and we should consider to adjust cycle length or green time.

5. Conclusions

We have presented a method of offset optimization based on queue length constraint for the saturated arterial, whose objective is to minimize queue length and avoid queue spillover over the whole duration of saturation. To avoid secondary queues, in algorithm development we assume that through flows from upstream are pulse discharge to downstream. VISSIM simulation results show the following: to ensure the operation parameters of the competitive phase to not be deteriorated, the offset which is calculated by our optimization method can effectively reduce the maximum queue length and prevent queuing spillover. In addition, sensitivity analysis of the model shows that, before the volume ratios exceed the threshold, the effectiveness by adjusting offset to reduce maximum queue length is less significant when volumes increase. The next step will be focused on an in-depth study of timing parameters of coordinated optimization strategy for the saturated arterial.

Acknowledgment

This work was supported by the National Science Foundation of China (Grant no. 50908100, Grant no. 51278520, and Grant no. 51278220) and China Postdoctoral Science Foundation funded project (Grant no. 20110491307). The authors thank anonymous reviewers for their valuable input and suggestions.

References

- [1] N. H. Gartner, J. D. C. Little, and H. Gabbay, "Optimization of traffic signal settings by mixed integer linear programming," *Transportation Science*, vol. 9, pp. 321–343, 1975.

- [2] N. H. Gartner, S. F. Assman, F. Lasaga, and D. L. Hou, "A multi-band approach to arterial traffic signal optimization," *Transportation Research Part B*, vol. 25, no. 1, pp. 55–74, 1991.
- [3] C. K. Wong and S. C. Wong, "A lane-based optimization method for minimizing delay at isolated signal-controlled junctions," *Journal of Mathematical Modelling and Algorithms*, vol. 2, no. 4, pp. 379–406, 2003.
- [4] J. Chang, B. Bertoli, and W. Xin, "New signal control optimization policy for oversaturated arterial systems," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, DC, USA, 2010.
- [5] A. L. Ghassan and F. B. Rahim, "Genetic algorithms for traffic signal control and queue management of oversaturated two-way arterials," *Transportation Research Record*, no. 1727, pp. 61–67, 2000.
- [6] D. C. Gazis and R. Herman, "The moving and phantom bottleneck," *Transportation Science*, vol. 26, no. 3, pp. 223–229, 1992.
- [7] G. F. Newell, "A moving bottleneck," *Transportation Research Part B*, vol. 32, no. 8, pp. 531–537, 1998.
- [8] Z. K. Khatib and G. A. Judd, "Control strategy for oversaturated signalized intersections," in *Proceedings of the Transportation Research Board 80th Annual Meeting*, Washington, DC, USA, 2001.
- [9] A. L. Ghassan and F. B. Rahim, "Development of traffic control and queue management procedures for oversaturated arterials," *Transportation Research Record*, no. 1603, pp. 119–127, 1997.
- [10] Y. T. Wu and C. H. Ho, "The development of Taiwan arterial traffic-adaptive signal control system and its field test: a Taiwan experience," *Journal of Advanced Transportation*, vol. 43, no. 4, pp. 455–480, 2009.
- [11] A. Sharma, D. M. Bullock, and J. A. Bonneson, "Input-output and hybrid techniques for real-time prediction of delay and maximum queue length at signalized intersections," *Transportation Research Record*, no. 2035, pp. 69–80, 2007.
- [12] J. Walmsley, "The practical implementation of SCOOT traffic control systems," *Traffic Engineering & Control*, vol. 23, no. 4, pp. 196–199, 1982.
- [13] T. H. Chang and J. T. Lin, "Optimal signal timing for an oversaturated intersection," *Transportation Research Part B*, vol. 34, no. 6, pp. 471–491, 2000.
- [14] E. B. Lieberman, J. Chang, and E. S. Prassas, "Formulation of real-time control policy for oversaturated arterials," *Transportation Research Record*, no. 1727, pp. 77–78, 2000.
- [15] W. H. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, China Communications Press, Beijing, China, 2012.
- [16] H. W. Guo, W. H. Wang, W. W. Guo, X. B. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [17] D. H. Wang, Y. M. Bie, X. M. Song, and H. S. Qi, "Maximum cycle length optimization method of signalized intersection," *Journal of Jilin University*, vol. 40, no. 1, pp. 148–151, 2010.
- [18] R. P. Roess and W. R. McShane, *Traffic Engineering*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2nd edition, 1990.
- [19] D. H. Wang, *Traffic Flow Theory*, China Communications Press, Beijing, China, 2002.

Research Article

An Improved Model for Headway-Based Bus Service Unreliability Prevention with Vehicle Load Capacity Constraint at Bus Stops

Weiya Chen,^{1,2} Chunhua Yang,¹ Fenling Feng,² and Zhiya Chen²

¹ School of Information Science and Engineering, Central South University, Changsha 410083, China

² School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

Correspondence should be addressed to Fenling Feng, ffl0731@163.com

Received 29 September 2012; Revised 22 November 2012; Accepted 27 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Weiya 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.

This paper presents an improved model for improving headway-based bus route service reliability at bus stops using real-time preventive operation control, taking into account dynamic interaction among random passenger demand, stochastic driving conditions of route segments, and vehicle load capacity constraint. In this model, the real-time information of passenger demand and vehicle operation is involved to predict the imminent unacceptable headway deviation, in the case of which some in-time preventive control strategies are deployed according to the given control rules. As a case study, a single fixed bus route with high-frequency services was simulated and different scenarios of real-time preventive operation control were performed. Headway adherence and average passenger wait time were used to measure bus service reliability. The results show that the improved model is closer to the real bus route service, and using real-time information to predict potential service unreliability and trigger in-time preventive control can reduce bus bunching and avoid big gap.

1. Introduction

Giving priority to the development of urban public transit is becoming the common view on reducing urban traffic jam and improving urban travel efficiency [1]. But it is not just a policy issue to attract more and more people to choose transit for travelling. A challenging problem faced by the government, researchers, and transit agencies is how to provide better transit service by using up-to-date technologies.

Reliability is one of the most important attributes of quality of transit service and always the top concerned issue for both passengers and transit agencies [2, 3]. From the perception of the passengers, service unreliability means more average wait time, which is

identified by Welding's assertion that the more regular service means the lower average wait time for potential passengers, especially on high-frequency bus routes with random passenger demand [4]. For transit agencies, service delays and disruptions have a real monetary cost in terms of lower utilization of vehicles and operators, which account for 3–5% of operating and vehicle costs by conservative estimate [5, 6]. In terms of the causes of unreliability, running time variability and passenger demand fluctuation are generally noted to be significant factors for service unreliability [7, 8]. Moreover, the initial headway irregularity, either at the beginning or the mid-route, will propagate downstream and this kind of propagation tends to worsen passenger load fluctuation and contributes to worse unreliability downstream [9, 10].

In order to improve bus service reliability, the provision of real-time information technologies has been recognized as a persuasive strategy in more and more countries. In recent days, real-time information technologies like automatic vehicle location (AVL) and automatic passenger count (APC) are used to monitor service and passenger demand, based on which different operation control strategies are adopted to control bus operation [11–15]. From the existing research, most of the control strategies were studied to restore service reliability after the occurrence of disruption [16, 17]. Considering real-time information collected by the advanced public transit system, however, it is expected to forewarn big headway deviation from the beginning of the route service and adopt some immediate preventive action to avoid potential service irregularity [18–21]. In addition, in a rich information environment, model-based intelligent vehicle systems would help drivers to achieve the effectiveness of the vehicle operation control [22, 23]. At this point, an effective model for decision-making using real-time information is the key issue for real-time bus operation control.

In [20], disregarding the vehicle load capacity constraint, we developed a simulation model for fixed-route transit service unreliability prevention based on AVL-APC data, whose effectiveness was extensively demonstrated on a circular bus route [21]. However, in the real world, vehicle load capacity is a very important factor affecting bus service reliability from the perspective of passengers. Based on the main framework of the model presented in [20], this paper develops an improved model for improving bus service reliability by using instantaneous prediction and preventive control strategies on high-frequency bus route, considering the dynamic interaction among random passenger demand, stochastic driving conditions of route segments and the vehicle load capacity constraint. In addition, it presents a set of decision-making rules for the potential intelligent dispatching module.

This rest of the paper is organized as follows. The next section presents the measures of headway-based transit service reliability. In Section 3, a vehicle-load-capacity-constrained preventive model is developed based on the main model framework in [20]. Section 4 presents a Monte Carlo simulation with the model embedded, which applies real-time preventive control approach to a simple fixed bus route with high-frequency services. Conclusions and future work are provided in Section 5.

2. Measures of Headway-Based Bus Service Reliability

2.1. Coefficient of Variation of Headways (C_{vh})

On high-frequency bus routes, passengers generally arrive at stops randomly disregarding the schedule. Therefore, headway adherence is the most important index for measuring service reliability. In the Transit Capacity and Quality of Service Manual (TCQSM) [24], the

Table 1: The level of service classified by C_{vh} for fixed-route bus service.

LOS	C_{vh}	$P(h_i > 0.5h)$	Comments
A	0.00–0.21	$\leq 1\%$	Service provided like clockwork
B	0.22–0.30	$\leq 10\%$	Vehicle slightly off headway
C	0.31–0.39	$\leq 20\%$	Vehicle often off headway
D	0.40–0.52	$\leq 33\%$	Irregular headways, with some bunching
E	0.53–0.74	$\leq 50\%$	Frequent bunching
F	≥ 0.75	$> 50\%$	Most vehicle bunched

Note: (a) The probability P means that a given transit vehicle's headway h_i will be off headway by more than one-half the scheduled headway h .

(b) Applies to routes with headways of 10 minutes or less.

Source: transit capacity and quality of service manual-2nd edition.

measure is based on the coefficient of variation of headways (C_{vh}) of transit vehicles serving a fixed route arriving at a stop and is calculated as follows:

$$C_{vh} = \frac{\text{Standard deviation of headway deviations}}{\text{mean schedule headway}}. \quad (2.1)$$

For fixed-route high-frequency bus services, the level of service (LOS) is classified according to the coefficient of variation of headways (see Table 1).

2.2. Average Passenger Wait Time

Welding's average passenger wait time is also widely used to measure headway-based transit service reliability, which is a function of the mean and variance of the headway [3]:

$$E(W) = \frac{1}{2} \left(E(H) + \frac{\text{Var}(H)}{E(H)} \right). \quad (2.2)$$

In (2.2), $E(W)$ is the expected passenger wait time, $E(H)$ is the expected headway, and $\text{Var}(H)$ is the variance of the headway. According to (2.2), when $E(H)$ is a constant, the smaller $\text{Var}(H)$ means the less average passenger wait time $E(W)$.

3. Model Development

In order to improve bus service reliability, static preventive control strategies include route planning, scheduling, and fleet and labor management [25]. On dynamic bus operation control, an in-time preventive strategy could be adopted to avoid big headway deviation before the imminent bus bunching or gap, which requires real-time bus service information for in-time decision making of operation control. In terms of headway-based preventive control, it is expected to keep $\text{Var}(H)$ in (2.2) close to zero as possible. Therefore, as done in [20], the prediction and forewarning of potential service irregularity using real-time information and the immediate action to prevent the headway from deviating too much is the underlying principle of this improved model.

Before the improved preventive model is formulated, we use a single fixed bus route for description. General assumptions are made and the notations used in this paper are listed below.

- (a) Vehicles arrive at each stop frequently (e.g., headways of 5 minutes), making it proper to assume random passenger arrivals at each stop. Moreover, the arrival rate and alighting fraction at a given stop do not change over the observed time period.
- (b) When a vehicle arrives at a stop, boarding and alighting of passengers can occur simultaneously (usually for vehicles with a front door for boarding and a rear door for alighting). The average boarding time and alighting time of each passenger are constant for simplicity.
- (c) The dwell time of a vehicle at a given stop is the maximum of the boarding time and the alighting time.
- (d) The running time of a vehicle in between two stops is random, but still predictable and controllable to some extent.
- (e) Considering the real-time information of passenger demand and vehicle operation, when a vehicle leaves the current bus stop, the dwell time and running time of its preceding vehicle, as well as its own leading headway and the on-board passenger number, are known for decision making of a preventive control strategy.
- (f) Considering vehicle load capacity, those passengers who are not able to take the current fully loaded bus have to wait for the next bus. Vehicles are not allowed to overtake each other under preventive control.

A set of variables will be used for describing the interaction between the vehicles and passengers on this fixed bus route.

H_{ik} = the leaving headway of the vehicle i departing from the stop k ;

R_{ik} = the running time of the vehicle i from stop $k - 1$ to stop k ;

D_{ik} = the dwell time of the vehicle i serving passengers' boarding and alighting at stop k ;

B_{ik} = the random number of passengers boarding the vehicle i at stop k ;

UB_{ik} = the number of passengers who are not able to get on the vehicle i at stop k due its capacity constraint and full loading;

A_{ik} = the random number of passengers alighting from the vehicle i at stop k ;

L_{ik} = the number of on-board passengers of the vehicle i when it departs from stop k ;

λ_k = the passenger arrival rate (passengers per minute) at stop k ;

ρ_k = the passenger alighting fraction of the on-board passengers at stop k ;

α = the average alighting time for each passenger;

β = the average boarding time for each passenger;

C = the vehicle load capacity.

3.1. Dynamic Interaction between Passengers and Vehicles

The route-level bus service system is a typical discrete dynamic system. The dynamic interaction between passengers and vehicles will be described based on the main model framework presented in [20]. The transit vehicles spend time on both traveling in between stops and dwelling at stops to serve passengers' boarding and alighting. The dwelling time of the vehicle i at stop k is decided by the maximum of the times of passengers' alighting and boarding. According to the assumption (b), D_{ik} will be decided by

$$D_{ik} = \max(\alpha \cdot A_{ik}, \beta \cdot B_{ik}). \quad (3.1)$$

Disregarding the vehicle load capacity C , A_{ik} and B_{ik} will be calculated or predicted by (3.2)

$$\begin{aligned} A_{ik} &= \rho_k \cdot L_{i, k-1}, \\ B_{ik} &= \lambda_k \cdot H_{ik}. \end{aligned} \quad (3.2)$$

Considering the constraint of the vehicle load capacity C , those passengers who are not able to get on the vehicle $i - 1$ at stop k will have to wait for the vehicle i . Thus we will have B_{ik} , UB_{ik} , and L_{ik} in (3.3) as

$$\begin{aligned} B_{ik} &= \min(C - (1 - \rho_k) \times L_{i, k-1}, \lambda_k \cdot H_{ik} + UB_{i-1, k}), \\ UB_{ik} &= \max(0, \lambda_k \cdot H_{ik} + UB_{i-1, k} - C + (1 - \rho_k) \times L_{i, k-1}), \\ L_{ik} &= L_{i, k-1} - A_{ik} + B_{ik}. \end{aligned} \quad (3.3)$$

When the vehicle i is departing from stop $k - 1$, the number of on-board passengers $L_{i, k-1}$ could be observed according to the real-time information. Thus, it is appropriate to predict the minimum D_{ik} based on the observed $L_{i, k-1}$, say, the vehicle i will have to dwell at stop k at least for a total alighting time ($\alpha \cdot \rho_k \cdot L_{i, k-1}$), which can be formulized as (3.4).

$$(D_{ik})_{\min} = \alpha \cdot \rho_k \cdot L_{i, k-1}. \quad (3.4)$$

The boarding time ($\beta \cdot \lambda_k \cdot H_{ik}$) will depend on the headway H_{ik} with given β and λ_k ; while the headway H_{ik} depends on the trajectories of transit vehicles. One step of iteration is as follows:

$$H_{ik} = H_{i, k-1} + \Delta R_{ik} + \Delta D_{ik}, \quad (3.5)$$

$$\Delta R_{ik} = R_{ik} - R_{i-1, k}, \quad (3.6)$$

$$\Delta D_{ik} = D_{ik} - D_{i-1, k}, \quad (3.7)$$

where ΔR_{ik} is the difference in running times between bus i and its preceding bus $i - 1$ when they arrive at stop k ; ΔD_{ik} is the difference in dwell times between bus i and its preceding bus $i - 1$ when they dwell at stop k .

The dynamic iterative calculation of the headway H_{ik} is as follows:

$$H_{ik} = H_{im} + \sum_{j=m+1}^k \Delta R_{ij} + \sum_{j=m+1}^k \Delta D_{ij}, \quad 0 \leq m \leq k - 1, \quad k > 1. \quad (3.8)$$

Particularly, let $m = 1$, the iteration will begin from the starting stop to the current stop k , then,

$$H_{ik} = H_{i1} + \sum_{j=2}^k \Delta R_{ij} + \sum_{j=2}^k \Delta D_{ij}, \quad k > 1. \quad (3.9)$$

From (3.8) and (3.9), the headway H_{ik} depends on the previous headways, the accumulated running time differences and dwell time differences, which present a theoretical proof for that the initial headway irregularity, either at the beginning or the mid-route, will propagate downstream. This kind of propagation is caused by the dynamic interactions between passengers and vehicles, and the worsen vehicle load fluctuation would contribute to the worse unreliability downstream, vice versa.

3.2. Real-Time Preventive Bus Operation Control

In a preventive model, the ideal headway H_{ik} will be always expected to be kept close to the scheduled headway H_i stop by stop ($k = 1, 2, 3, \dots, N$) along the bus route. Mathematically, H_{ik} is kept on the interval decided by H_i plus or minus a permitted deviation ε ($\varepsilon > 0$) due to the stochastic operation environment as

$$H_i - \varepsilon \leq H_{ik} \leq H_i + \varepsilon, \quad \varepsilon > 0. \quad (3.10)$$

With the given vehicle load capacity and the initial headway, the headway H_{ik} will mainly depend on the total running time difference and the total dwell time difference in between the first stop and the current stop k . When it comes to the operation control strategies, two options of control strategies will be decided. One option is bus holding strategy for controlling the vehicle dwell time; the other option is bus speed adjusting for controlling the vehicle running time. The bus speed adjusting strategy will be preferred on high-frequency bus route for two aspects of consideration. On one hand, the bus holding strategy works well for improving service reliability, but it has some disadvantages: wasting time of the on-board passengers, occupying the pick-up location at the bus station, delaying the arrival at the destination, and so forth [24]. On the other hand, the bus speed adjusting is acceptable and operational, which is benefited from the real-time traffic information technologies. Determining how to use the two strategies will be the key issue for the preventive control problem.

To illustrate this problem clearly, a simple example bus route is used (see Figure 1). On this single fixed bus route, the bus i is leaving current stop $k - 1$ to stop k ; with the

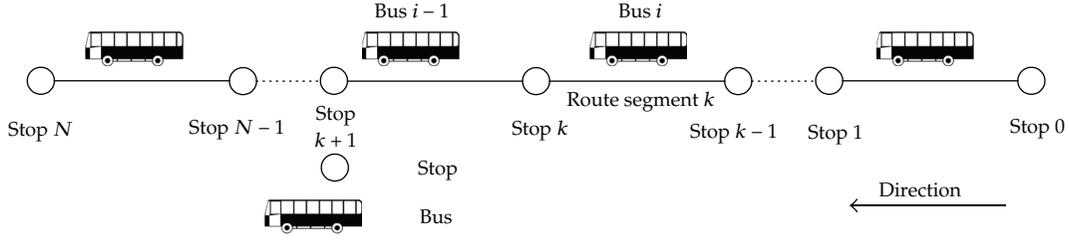


Figure 1: Example bus route.

observed $H_{i,k-1}$, $L_{i,k-1}$, $R_{i-1,k}$, and $D_{i-1,k}$ (according to the assumption(e)), we need to predict and suggest a reasonable running time R'_{ik} and D'_{ik} for bus i to keep an acceptable H'_{ik} . Determining an adaptive control strategy will follow the steps below.

Step 1. To decide the predicted dwell time \tilde{D}_{ik} . As discussed above, a minimum dwell time $(D_{ik})_{\min}$ is the alighting time and can be calculated by (3.4); and only if the boarding time $(\beta \cdot \lambda_k \cdot H_{ik})$ is larger than the alighting time, the dwell time will equal to the boarding time. Here, we use a dummy variable φ to determine the predicted dwell time \tilde{D}_{ik} , which is calculated as

$$\tilde{D}_{ik} = \varphi \cdot \alpha \cdot \rho_k \cdot L_{i,k-1} + (1 - \varphi) \cdot \beta \cdot \lambda_k \cdot \bar{H}_i, \quad (3.11)$$

where

$$\varphi = \begin{cases} 1 & \text{for } \alpha \cdot \rho_k \cdot L_{i,k-1} \geq \beta \cdot \lambda_k \cdot \bar{H}_i, \\ 0 & \text{otherwise.} \end{cases} \quad (3.12)$$

\bar{H}_i is the expected headway of the vehicle i .

As discussed above, the vehicle speed control strategy will be firstly adopted and then will be the holding strategy. Therefore, the next step is to decide a reasonable running time.

Step 2. To decide a reasonable running time R'_{ik} . A reasonable running time is decided jointly by the feasible running time and the acceptable running time. With the transfer of (3.5), an acceptable running time \bar{R}_{ik} is given as follows:

$$\begin{aligned} (\bar{H}_i - \varepsilon - H_{i,k-1} + R_{i-1,k} + D_{i-1,k}) - \tilde{D}_{ik} &\leq \bar{R}_{ik} \\ &\leq (\bar{H}_i + \varepsilon - H_{i,k-1} + R_{i-1,k} + D_{i-1,k}) + \tilde{D}_{ik}. \end{aligned} \quad (3.13)$$

A feasible running time \overleftarrow{R}_{ik} is decided by the real-time traffic conditions in between stop $k-1$ and stop k . Along with the development of intelligent transportation technologies, it will be easy to get the real-time road traffic information, such as current traffic volume and speed. Here we still use a prediction way to decide the feasible running time \overleftarrow{R}_{ik} . It

is acceptable to predict it based on $R_{i-1,k}$ or the statistical mean running time \overline{R}_k and the standard deviation σ , respectively,

$$R_{i-1,k} \cdot (1 - \delta_1) \leq \overleftarrow{R}_{ik} \leq R_{i-1,k} \cdot (1 + \delta_2), \quad (3.14)$$

$$\overline{R}_k - x \cdot \sigma \leq \overleftarrow{R}_{ik} \leq \overline{R}_k + y \cdot \sigma. \quad (3.15)$$

In (3.14), δ_1 and δ_2 are nonnegative percentile fractions. In (3.15), x and y are nonnegative parameters used to adjust the deviation and reflect the real-time traffic conditions. Considering the combination of (3.13) and (3.14) or (and) (3.15), a reasonable R'_{ik} will be located on the interval decided jointly by \overrightarrow{R}_{ik} and \overleftarrow{R}_{ik} :

$$R'_{ik} \in \left[\left(\overrightarrow{R}_{ik} \right)_{\min}, \left(\overrightarrow{R}_{ik} \right)_{\max} \right] \cap \left[\left(\overleftarrow{R}_{ik} \right)_{\min}, \left(\overleftarrow{R}_{ik} \right)_{\max} \right]. \quad (3.16)$$

Those values over the interval in (3.16) are the suggested running time for the vehicle i leaving for stop k . Theoretically, the interval is possible to be void, which means the acceptable \overrightarrow{R}_{ik} can not be satisfied, either because the minimum \overrightarrow{R}_{ik} is bigger than the maximum feasible running time \overleftarrow{R}_{ik} or because the maximum \overrightarrow{R}_{ik} is smaller than the minimum \overleftarrow{R}_{ik} . In practice, the former reason is impossible due to an infinite \overleftarrow{R}_{ik} ; then the bus will be suggested to run at any \overrightarrow{R}_{ik} . The latter situation usually means the bad traffic conditions, and bus i will be requested to run at the maximum feasible speed to prevent the gap. By using the theoretical analysis, we make some decision-making rules for the reasonable running time as follows:

- (1) If $\left(\overrightarrow{R}_{ik} \right)_{\max} \leq \left(\overleftarrow{R}_{ik} \right)_{\min}$, then $R'_{ik} = \left(\overleftarrow{R}_{ik} \right)_{\min}$;
- (2) If $\left(\overrightarrow{R}_{ik} \right)_{\min} \leq \left(\overleftarrow{R}_{ik} \right)_{\min}$ and $\left(\overrightarrow{R}_{ik} \right)_{\max} \geq \left(\overleftarrow{R}_{ik} \right)_{\min}$ and $\left(\overrightarrow{R}_{ik} \right)_{\max} \leq \left(\overleftarrow{R}_{ik} \right)_{\max}$, then $R'_{ik} \in \left[\left(\overleftarrow{R}_{ik} \right)_{\min}, \left(\overrightarrow{R}_{ik} \right)_{\max} \right]$;
- (3) If $\left(\overrightarrow{R}_{ik} \right)_{\min} \leq \left(\overleftarrow{R}_{ik} \right)_{\min}$ and $\left(\overrightarrow{R}_{ik} \right)_{\max} \geq \left(\overleftarrow{R}_{ik} \right)_{\max}$, then $R'_{ik} \in \left[\left(\overleftarrow{R}_{ik} \right)_{\min}, \left(\overleftarrow{R}_{ik} \right)_{\max} \right]$;
- (4) If $\left(\overrightarrow{R}_{ik} \right)_{\min} \geq \left(\overleftarrow{R}_{ik} \right)_{\min}$ and $\left(\overrightarrow{R}_{ik} \right)_{\max} \leq \left(\overleftarrow{R}_{ik} \right)_{\max}$, then $R'_{ik} \in \left[\left(\overrightarrow{R}_{ik} \right)_{\min}, \left(\overrightarrow{R}_{ik} \right)_{\max} \right]$;
- (5) If $\left(\overrightarrow{R}_{ik} \right)_{\min} \geq \left(\overleftarrow{R}_{ik} \right)_{\min}$ and $\left(\overrightarrow{R}_{ik} \right)_{\min} \leq \left(\overleftarrow{R}_{ik} \right)_{\max}$ and $\left(\overrightarrow{R}_{ik} \right)_{\max} \geq \left(\overleftarrow{R}_{ik} \right)_{\max}$, then $R'_{ik} \in \left[\left(\overrightarrow{R}_{ik} \right)_{\min}, \left(\overleftarrow{R}_{ik} \right)_{\max} \right]$;
- (6) If $\left(\overrightarrow{R}_{ik} \right)_{\min} \leq \left(\overleftarrow{R}_{ik} \right)_{\max}$, then $R'_{ik} = \left(\overrightarrow{R}_{ik} \right)_{\min}$.

4. Simulation and Results

4.1. Parameters Description

In order to clarify this model, the example bus route in Figure 1 is also used for the case study. This bus route has 11 stops which are scattered about in a fairly uniformity of running time, with buses being dispatched from stop 0 to stop 10 at 5-minute headway ($E(H) = 5$). The vehicle load capacity is 80 persons ($C = 80$).

Table 2: Example of bus route parameters.

Stop	$\lambda_k /$ Passengers/min	$\rho_k /$ %	$E(R_k) /$ min	$E(H) /$ min
0	1.0	0.0	—	5.0
1	0.75	0.0	3	5.0
2	0.75	0.1	3	5.0
3	2.0	0.2	3	5.0
4	2.0	0.25	3	5.0
5	3.0	0.6	3	5.0
6	1.5	0.25	3	5.0
7	0.75	0.25	3	5.0
8	0.5	0.5	3	5.0
9	0.1	0.1	3	5.0
10	0.0	1.0	3	5.0

Each bus is expected to run in between two successive stops in 3 minutes ($E(R_k) = 3$), leading to a total expected running time of 30 minutes from stop 0 to stop 10 in one direction. The dwell time at each stop depends on the passenger demand. On high-frequency bus route, usually passenger arrival at each stop is a Poisson process and the number of passengers alighting at each stop is subject to a binomial probability distribution based on the current passenger load on the vehicle [6, 13]. Here the passenger demand is characterized with a Poisson passenger arrival rate λ_k and a binomial probability of passenger alighting fraction ρ_k . In addition, the boarding time and the alighting time per passenger are assumed to be 3.6 seconds and 1.8 seconds ($\alpha = 0.06, \beta = 0.03$). Table 2 gives the parameters of the operating conditions of the example bus route.

4.2. Simulation Scenarios

For illustrative purposes, in a single simulation period 30 buses from stop 0 to stop 10 are observed and a total of 20 simulation runs are performed for four cases.

Case 1. Running time R_{ik} is fluctuated with $\delta_1 = \delta_2 = 0.4$ and $\varepsilon = 60$ seconds.

Case 2. Running time R_{ik} is fluctuated with $\delta_1 = \delta_2 = 0.4$ and $\varepsilon = 90$ seconds.

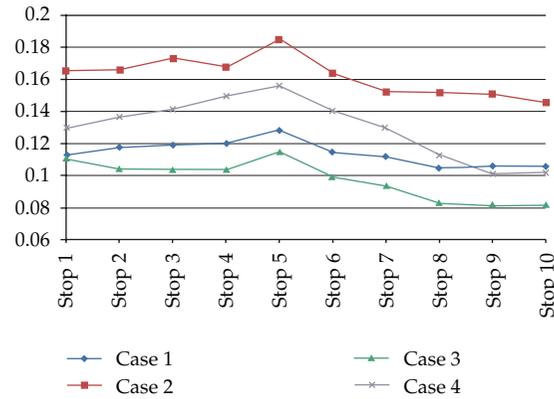
Case 3. Running time R_{ik} is fluctuated with $\delta_1 = \delta_2 = 0.2$ and $\varepsilon = 60$ seconds.

Case 4. Running time R_{ik} is fluctuated with $\delta_1 = \delta_2 = 0.2$ and $\varepsilon = 90$ seconds.

The initial conditions for each single simulation run includes: each bus leaves stop 0 precisely at its given headway, and the first bus is supposed to run in between two consecutive stops precisely with 3 minutes. The preventive strategies will be triggered by checking the real-time conditions following Steps 1 and 2 in Section 3.2.

Table 3: The observed average headway, standard deviation of headways, and coefficient of variation of headways.

Scenarios	Stop 1	Stop 2	Stop 3	Stop 4	Stop 5	Stop 6	Stop 7	Stop 8	Stop 9	Stop 10
Case 1 Avg hdw	5.0645	5.0964	5.1042	5.1091	5.1340	5.1497	5.1616	5.1694	5.1698	5.1568
Case 1 Std hdw	0.5636	0.5900	0.5965	0.6000	0.6414	0.5741	0.5603	0.5238	0.5305	0.5290
Case 1 C_{vh}	0.1127	0.1180	0.1193	0.1200	0.1283	0.1148	0.1121	0.1048	0.1061	0.1058
Case 2 Avg hdw	5.1093	5.1919	5.2702	5.3087	5.3304	5.3670	5.4001	5.4170	5.4103	5.4173
Case 2 Std hdw	0.8262	0.8295	0.8644	0.8379	0.9244	0.8189	0.7609	0.7592	0.7541	0.7280
Case 2 C_{vh}	0.1652	0.1659	0.1729	0.1676	0.1849	0.1638	0.1522	0.1518	0.1508	0.1456
Case 3 Avg hdw	5.1290	5.1933	5.2185	5.2655	5.2988	5.3456	5.3819	5.4005	5.3938	5.3968
Case 3 Std hdw	0.5529	0.5221	0.5186	0.5188	0.5745	0.4960	0.4676	0.4147	0.4073	0.4078
Case 3 C_{vh}	0.1106	0.1044	0.1037	0.1038	0.1149	0.0992	0.0935	0.0829	0.0815	0.0816
Case 4 Avg hdw	5.1831	5.3577	5.4580	5.5623	5.6341	5.7012	5.7574	5.7952	5.8196	5.8311
Case 4 Std hdw	0.6506	0.6836	0.7083	0.7497	0.7800	0.7011	0.6483	0.5651	0.5078	0.5089
Case 4 C_{vh}	0.1301	0.1367	0.1417	0.1499	0.1560	0.1402	0.1297	0.1130	0.1016	0.1018

**Figure 2:** Stop-level changing of coefficient of variance of headways.

4.3. Results and Discussion

The two metrics described in Section 2 are used to measure service reliability. The first one is the observed average headway and the coefficient of variation of headways (C_{vh}) at each stop. Table 3 gives the observed average headway (Avg hdw), the standard deviation of headways (std hdw), and the coefficient of variation of headways (C_{vh}) for each case.

According to Table 3, the average headways in four cases are below 6 minutes and the coefficients of variation of headways are below 0.2. It also shows that, in all four cases, the biggest coefficients of variation of headways appear at stop 5 due to the highest probability of passenger boarding and alighting. The route-level changing of the coefficient of variation of headways is further shown in Figure 2. Compared with the results in [26], these results indicate good service unreliability prevention.

Comparing Cases 1 with 2, the running time fluctuation is the same with 40% of the scheduled running time, but the permitted headway deviation is different. From the simulation as shown in Figure 2, the coefficients of variation of headways at each stop in Case 1 are smaller than those of Case 2, which indicates that better service unreliability prevention

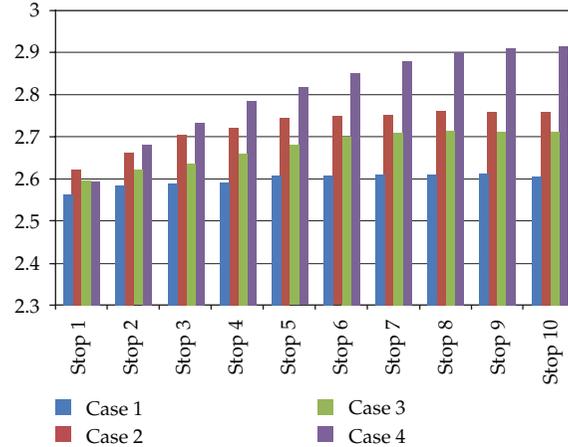


Figure 3: Average passenger wait time at each stop.

has been achieved in Case 1. The same situation appears when comparing Cases 3 with 4. These results suggest that the tougher request for keeping headway adherence promises better bus operation control. The practical reason might be that a bus driver may pay more attention to improving bus service reliability under strict performance evaluation.

Comparing Cases 1 with 3, the permitted headway deviation is the same with one minute, but the running time fluctuation is different. In Case 3, the lower limit of the running time fluctuation is half of that in Case 1. The coefficients of variation of headways at each stop in Case 3 are smaller than those in Case 1. The same situation also appears when comparing Cases 2 with 4. The results suggest that, requested by keeping the same permitted headway deviation, the smaller lower limit of the running time fluctuation is more likely to result in better headway regularity.

The other metric is average passenger wait time. Figure 3 shows the average passenger wait time at each stop in four cases.

As shown in Figure 3, the average passenger wait time at each stop is controlled at a level of less than 3 minutes in all four cases, but it still has a growing tendency. The best preventive effectiveness appears in Case 1, the worst in Case 4. Compared with the route-level changing of the coefficient of variation of headways, the growing tendency of the average passenger wait time is an interesting finding. This is noteworthy in that the intuition from the previous studies suggests that controlling one vehicle only by referring to its preceding vehicle might have a time lag of the travel time.

5. Conclusions and Future Works

This paper presents an improved preventive model for improving headway-based bus service reliability at bus stops, considering the constraint of the vehicle load capacity, as well as the dynamic nature of the random passenger demand and the stochastic road traffic conditions. Standing on this point, this model represents a more realistic approach to characterize the route-level bus service system and to utilize real-time information in a real-time decision-making context. As demonstrated in the simulation example, using some

probability distributions of the running time and the constant passenger boarding time and alighting time, the model and the preventive control strategy are effective.

By using the simulation-based methodology, the model presented in this paper mainly presents a way of how to follow the steps and the decision-making rules to utilize the real-time information to make better decision. Nonetheless, a more practical decision-making supporting system or a more intelligent decision-making module will be more powerful and useful in the real application. In this sense, this model can be directly embedded in the advanced public transit system, assisting the dispatching center and the operators to utilize the real-time information including vehicle running time, speed, dwell time, passenger demand, and other unreliability causes to instantaneously predict and forewarn big headway deviation of high-frequency bus routes, based on which better decisions can be made regarding the use of preventive control strategies. This job has become our ongoing work, and a real bus route will be chosen in the city of Zhuzhou in China for testing the real effectiveness. With the real application and test, the improvements of this proposed methodology could be promised. More broadly, in terms of the preventive control strategies, the route-level multivehicle collaborative control based on synchronous simulation optimization is another important future work.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) under Grants nos. 61203162 and 71171200. This work was also supported in part by the Postdoctoral Science Foundation of Central South University, the Freedom Explore Program of Central South University (Grant no. 721500036), and the Fundamental Research Funds for the Central Universities (Grant no. 2010QZZD021).

References

- [1] M. Xu, A. Ceder, Z. Gao, and W. Guan, "Mass transit systems of Beijing: governance evolution and analysis," *Transportation*, vol. 37, no. 5, pp. 709–729, 2010.
- [2] Y. Yin, W. H. K. Lam, and M. A. Miller, "A simulation-based reliability assessment approach for congested transit network," *Journal of Advanced Transportation*, vol. 38, no. 1, pp. 27–44, 2003.
- [3] C. F. Daganzo and J. Pilachowski, "Reducing bunching with bus-to-bus cooperation," *Transportation Research B*, vol. 45, no. 1, pp. 267–277, 2011.
- [4] P. I. Welding, "The instability of a close-interval service," *Operation Research Quarterly*, vol. 8, no. 3, pp. 133–148, 1957.
- [5] M. H. Abkowitz, R. Slavin, L. Waksman, and N. W. English, *Transit Service Reliability*, United States Department of Transportation, Washington, DC, USA, 1978.
- [6] M. D. Hickman, "An analytic stochastic model for the transit vehicle holding problem," *Transportation Science*, vol. 35, no. 3, pp. 215–237, 2001.
- [7] M. H. Abkowitz and I. Enelstin, "Factors affecting run time on transit routes," *Transportation Research A*, vol. 17, no. 2, pp. 107–113, 1983.
- [8] L. C. Cham, *Understanding Bus Service Reliability: A Practical Framework Using AVL/APC Data*, Massachusetts Institute of Technology, Cambridge, Mass, USA, 2006.
- [9] J. G. Strathman, T. J. Kimple, and S. Callas, *Headway Deviation EffectS on BuS PaSSenger LoadS: AnalySiS of Tri-Met'S Archived AVL-APC Data*, Portland State University, Portland, Ore, USA, 2003.
- [10] Z. R. Peng, E. D. Lynde, and W. Y. Chen, *Improving Service Restoration Using Automatic Vehicle Location,,* University of Wisconsin-Milwaukee, Milwaukee, Wis, USA, 2008.
- [11] P. G. Furth, B. J. Hemily, H. J. Muller, and J. G. Strathman, *Uses of Archived AVL-APC Data to Improve Transit Performance and Management: Review and Potential*, Transportation Research Board, Washington, DC, USA, 2003.

- [12] A. Shalaby and A. Farhan, "Prediction model of bus arrival and departure times using AVL and APC data," *Journal of Public Transportation*, vol. 7, no. 1, pp. 41–61, 2004.
- [13] I. E. Moses, *A Transit Route Simulator for the Evaluation of Control Strategies Using Automatically Collected Data*, Massachusetts Institute of Technology, Cambridge, Mass, USA, 2005.
- [14] C. A. Pangilinan, N. Wilson, and A. Moore, *Bus Supervision Deployment Strategies and the Use of Real-Time AVL for Improved Bus Service Reliability*, Transportation Research Board, Washington, DC, USA, 2008.
- [15] A. M. El-Geneidy, J. Horning, and K. J. Krizer, *Analyzing Transit Service Reliability Using Detailed Data from Automatic Vehicular Locator Systems*, Transportation Research Board, Washington, DC, USA, 2008.
- [16] J. G. Strathman, T. J. Kimpel, and K. J. Dueker, "Bus transit operations control: review and an experiment involving Tri-Met's automated bus dispatch system," *Journal of Public Transportation*, vol. 41, no. 1, pp. 1–26, 2001.
- [17] A. Sun and M. Hickman, "The holding problem at multiple holding stations," *Lecture Notes in Economics and Mathematical Systems*, vol. 600, pp. 339–359, 2008.
- [18] C. F. Daganzo, "A headway-based approach to eliminate bus bunching: systematic analysis and comparisons," *Transportation Research B*, vol. 43, no. 10, pp. 913–921, 2009.
- [19] Y. Hadas and A. Ceder, "Optimal coordination of public-transit vehicles using operational tactics examined by simulation," *Transportation Research C*, vol. 18, no. 6, pp. 879–895, 2010.
- [20] W. Y. Chen and Z. Y. Chen, "A simulation model for transit service unreliability prevention based on AVL-APC data," in *Proceedings of the International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '09)*, vol. 2, pp. 184–188, April 2009.
- [21] W. Y. Chen and Z. Y. Chen, "A bus-following model for preventing service unreliability on a circular bus route," in *Proceedings of the 2nd International Conference on Intelligent Computing Technology and Automation (ICICTA '09)*, vol. 3, pp. 425–428, October 2009.
- [22] W. H. Wang, H. W. Guo, H. Bubb, and K. Ikeuchi, "Numerical simulation and analysis procedure for model-based digital driving dependability in intelligent transport system," *KSCE Journal of Civil Engineering*, vol. 15, no. 5, pp. 891–898, 2011.
- [23] W. H. Wang, *Vehicle's Man-Machine Interaction Safety and Driver ASSiStance*, China Communications Press, Beijing, China, 2012.
- [24] KITTELSON & ASSOCIATES, Inc. and KFH Group, Inc., *Transit Capacity and Quality of Service Manual*, Transportation Research Board, Washington, DC, USA, 2nd edition, 2003.
- [25] M. A. Turnquist, "Strategies for improving reliability of bus transit service," *Transportation Research Record*, vol. 818, pp. 7–13, 1981.
- [26] W. Y. Chen and Z. Y. Chen, "Service reliability analysis of high frequency transit using stochastic simulation," *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 5, pp. 130–134, 2009.

Research Article

Forecasting Cohesionless Soil Highway Slope Displacement Using Modular Neural Network

**Yanyan Chen, Shuwei Wang, Ning Chen,
Xueqin Long, and Xiru Tang**

Transportation Research Center, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Shuwei Wang, wangshuwei1010@foxmail.com

Received 5 September 2012; Revised 16 November 2012; Accepted 21 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Yanyan 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.

The highway slope failures are triggered by the rainfall, namely, to create the disaster. However, forecasting the failure of highway slope is difficult because of nonlinear time dependency and seasonal effects, which affect the slope displacements. Starting from the artificial neural networks (ANNs) since the mid-1990s, an effective means is suggested to judge the stability of slope by forecasting the slope displacement in the future based on the monitoring data. In order to solve the problem of forecasting the highway slope displacement, a displacement time series forecasting model of cohesionless soil highway slope is given firstly, and then modular neural network (MNN) is used to train it. With the randomness of rainfall information, the membership function based on distance measurement is constructed; after that, a fuzzy discrimination method to sample data is adopted to realize online subnets selection to improve the self-adapting ability of artificial neural networks (ANNs). The experiment on the sample data of Beijing city's highway slope demonstrates that this model is superior to others in accuracy and adaptability.

1. Introduction

Forecasting the slope displacement accurately in the future becomes a very important step to judge the stability of slope. Forecasting the highway slope displacement in the future is a difficult task which involves evaluation of a large number of interrelated variables and factors and is hardly accounted for a model. Factors affecting slope stability can be divided into two main types: internal factors and external inducements. Internal factors include rock type, joints, property, and structure. External inducements include rainfalls, earthquakes, and human factors. Generally speaking, the slope failures are directly caused by the external inducements, and internal factors are the fundamentals of it. Also, most of these factors have obvious randomness, fuzziness, and variability. All factors interleaving together make the

influence on the stability of slope, so evaluation criterion which is used to judge the stability of slope should be easily measured and can make clearly manifestations of the change to slope stability.

Displacement is the external expression of deformation process to mass movements of soil and rock. In the past, displacement had been used to compute the prealert, alert, and emergency phases of large rock slides; this method had been validated by collecting and analysing literature data for historical rock slope failures [1]. The forecasting model of slope displacement includes artificial neural networks (ANNs) [2–7], grey theory [8–10], whole-region method of chaotic time series to slope deformation prediction [11]. The engineering properties of slope exhibit varied and uncertain behaviors, for the complex and imprecise physical processes associated with the formation of soil and rock [12]. So displacement considered as the external expression of slope failure shows evident nonlinear and uncertain characteristics. In order to cope with the complexity of behaviors and the spatial variability of displacement, traditional forms of forecasting models are justifiably simplified. Artificial neural networks (ANNs) refer to a class of artificial intelligence which attempt to mimic the behaviors of the human brain and nervous system, based on the data alone to determine the structure and parameters of the model. It is well suited to complex model problems where the relationship between the model variables is unknown. So, models based on ANNs have been successfully applied on virtually every problem in slope stability. However, rainfalls have important influence on the slope displacement [13–15], the forecasting displacement models of highway slope based on ANNs are focus on the time series variation of displacement, but neglected it.

Modular neural network (MNN) can solve large-scale real-world problems by dividing a problem into smaller and simpler subproblems, assigning a network module to learn each of the subproblems, and recombining the individual modules into a solution to the original problem. In this paper, using rainfall information, the membership function based on distance measurement is constructed, firstly, then a fuzzy discrimination method of sample data is adopted on on-line sub-nets selection in order to improve the self-adapting ability of artificial neural networks (ANNs). The experiment with the sample data of Beijing city's highway slope demonstrates that this model is superior to others in accuracy and adaptability.

2. Study the Relationship between Displacement and Rainfall

In order to obtain an accurate forecasting model based on MNN, a theoretical analysis of the relationship between displacement and rainfall has been made. The most common failure mode of cohesionless soil highway slope is landslide, the geological conditions is high permeability weathered layer or slope deposits which covers low hydraulic conductivity rock stratum, and the interface of them is often the potential slip surface. The result of analytic solution from Iverson [16] is used to describe the transient seepage due to rainfall. Then, an elasto-plastic analysis was used to analyze the unstable slope based on Mohr-Coulomb yield criterion and associated flow rule.

Figure 1 shows the force diagram of cohesionless soil highway slope in plane strain condition based on Mohr-Coulomb yield criterion. The assumptions are as follows:

- (i) depth of impervious surface and initial groundwater level is denoted by d and d_w , both of them are parallel to the slope surface;

- (ii) the homogeneous isotropic slide body is elastic-perfectly plastic material, strictly observing the Mohr-Coulomb yield and its flow rule;
- (iii) the source of groundwater recharge is only the rainfall; during this process without considering the loss such as evaporation, the soil above the groundwater level is fully saturated soil. In other words, all sliding soil masses have the same weight density (denoted by γ_{sat});
- (iv) ignoring the initial plastic strain of sliding soil mass before raining, at first, the initial effective stress parallel to the slope surface (denoted by σ'_{x0}) is proportional to initial effective stress perpendicular to the slope surface (denoted by σ'_{y0}) and is calculated using

$$\sigma'_{x0} = k_0 \sigma'_{y0}, \quad (2.1)$$

where k_0 is coefficient of lateral earth pressure. To elastic material, k_0 can be calculated by poisson ratio (denoted by μ) using

$$k_0 = \frac{\mu}{1 - \mu}. \quad (2.2)$$

In additional, it is assumed that intensity rainfall has a constant value during the entire process. Based on the analytical solution of Iverson, pore-water pressure with depth Y at time t (denoted by $\phi(Y, t)$) can be calculated by

$$\begin{aligned} \phi(Y, t) = & (Y - d_w) \cos^2 \alpha \\ & + 2 \frac{I}{K} \sqrt{D_1 t} \sum_{m=1}^{\infty} \left\{ i \operatorname{erfc} \left[\frac{(2m-1)d - (d-Y)}{2\sqrt{D_1 t}} \right] + i \operatorname{erfc} \left[\frac{(2m-1)d + (d-Y)}{2\sqrt{D_1 t}} \right] \right\} \\ & - 2 \frac{I}{K} H(t-T) \sqrt{D_1(t-T)} \sum_{m=1}^{\infty} \left\{ i \operatorname{erfc} \left[\frac{(2m-1)d - (d-Y)}{2\sqrt{D_1(t-T)}} \right] \right. \\ & \left. + i \operatorname{erfc} \left[\frac{(2m-1)d + (d-Y)}{2\sqrt{D_1(t-T)}} \right] \right\}, \end{aligned} \quad (2.3)$$

where K is osmotic coefficient of fully saturated soil, $H(\eta)$ is Heaviside step function, α is slope angle, D_0 is saturated hydraulic conductivity, and D_1 can be calculated using

$$D_1 = D_0 \cos 2\alpha. \quad (2.4)$$

Function $i \operatorname{erfc}(\eta)$ is given by

$$i \operatorname{erfc}(\eta) = \frac{1}{\sqrt{\pi}} \exp(-\eta^2) - \eta \operatorname{erfc}(\eta), \quad (2.5)$$

where $\operatorname{erfc}(\eta)$ is complementary error function.

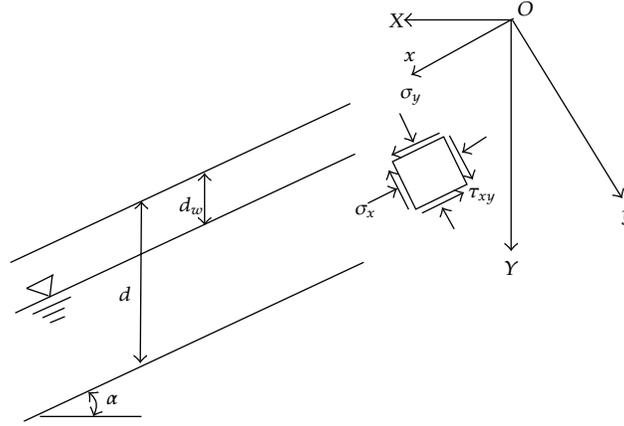


Figure 1: Force diagram of cohesionless soil highway slope.

Otherwise, $\phi(Y, t)$ should meet the requirement of

$$\phi(Y, t) \leq Y \cos^2 \alpha. \quad (2.6)$$

Hence, the factor of slope safety (defined by F_s) with depth Y at time t can be calculated using

$$F_s(Y, t) = \frac{\tan \varphi'}{\tan \alpha} + \frac{c' - \phi(Y, t) \gamma_w \tan \varphi'}{\gamma_{\text{sat}} Y \sin \alpha \cos \alpha}, \quad (2.7)$$

where φ' and c' are effective angles of internal friction and effective cohesion of soil, and γ_w is weight density of groundwater.

As shown in Figure 1, total normal stress (denoted by σ_Y) and total shear stress (denoted by τ_Y) with depth Y parallel to the slope surface are calculated using

$$\begin{aligned} \sigma_Y &= \gamma_{\text{sat}} Y \cos^2 \alpha, \\ \tau_{xy} &= \gamma_{\text{sat}} Y \cos \alpha \sin \alpha. \end{aligned} \quad (2.8)$$

Pore-water pressure with depth Y at time t (denoted by $u(Y, t)$) can be calculated using (2.9).

$$u(Y, t) = \phi(Y, t) \gamma_w. \quad (2.9)$$

Effective normal stresses which are parallel to the slope surface (denoted by $\sigma'_x(Y, t)$) and are perpendicular to the slope surface (denoted by $\sigma'_y(Y, t)$) are given by

$$\begin{aligned} \sigma'_x(Y, t) &= k_0 \sigma'_{y0} - \phi(Y, t) \gamma_w, \\ \sigma'_y(Y, t) &= \sigma_y - \phi(Y, t) \gamma_w. \end{aligned} \quad (2.10)$$

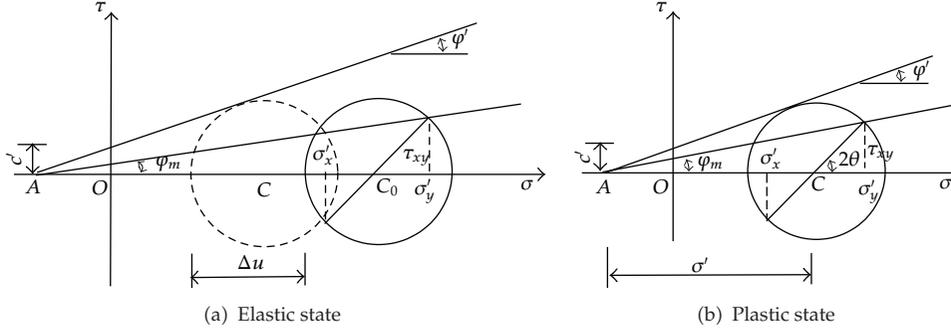


Figure 2: Stress state.

The relationship between Mohr's Circle of Stress and shear strength on site is shown in Figure 2.

Variable $\tan \varphi_m$ is defined

$$\tan \varphi_m = \frac{\tau_{xy}}{\sigma'_y + c' \cot \varphi'} \quad (2.11)$$

As shown in Figure 2(a), (2.12) is satisfied:

$$\frac{\sqrt{(\sigma'_y - \sigma'_x)^2 + 4\tau_{xy}^2}}{(\sigma'_y + \sigma'_x) + 2c' \cot \varphi'} \leq \sin \varphi' \quad (2.12)$$

Figure 2(a) shows that the increased pore-water pressure after the rainfall will shift the Mohr's Circle of Stress to the left till it is tangent to the curve of shear strength, and it becomes yield at this site, then inequation (2.12) will become equation. However, the yield surface due to yield point does not parallel the bottom plane of slope at this moment (Figure 2(b)); direction of principal stress axis will change with increased pore-water pressure until become parallel to the bottom plane of slope. During this period, stress of plastic zone can be calculated using

$$\begin{aligned} \sigma'_x &= \sigma' (1 - \sin \varphi' \cos 2\theta) - c' \cot \varphi', \\ \sigma'_y &= \sigma' (1 + \sin \varphi' \cos 2\theta) - c' \cot \varphi', \\ \tau_{xy} &= \sigma' \sin \varphi' \sin 2\theta, \end{aligned} \quad (2.13)$$

where θ is the angle of the maximum principal stress from y -axis, and it is given as

$$\tan \varphi_m = \frac{\sin \varphi' \sin 2\theta}{1 + \sin \varphi' \cos 2\theta} \quad (2.14)$$

Under plane strain conditions, Mohr-Coulomb yield criterion can be expressed as F , calculated using

$$F = \sqrt{\frac{1}{4}(\sigma'_x - \sigma'_y)^2 + \tau_{xy}^2} + \frac{1}{2}(\sigma'_x + \sigma'_y) \sin \varphi' - c' \cos \varphi' = 0. \quad (2.15)$$

Using the flow rule given by it, strain increment along the critical slip surface has a translational movement, also the value of it is zero, namely, $d\varepsilon_x = 0$; the relations between stress and strain increment are taken to be as

$$\begin{aligned} d\sigma'_x &= M_{12}d\varepsilon_y + M_{13}d\gamma_{xy}, \\ d\sigma'_y &= M_{22}d\varepsilon_y + M_{23}d\gamma_{xy}, \\ d\tau_{xy} &= M_{32}d\varepsilon_y + M_{33}d\gamma_{xy}, \end{aligned} \quad (2.16)$$

in which

$$\begin{aligned} M_{12} &= \frac{AB}{D} + \lambda, \\ M_{13} &= \frac{CB}{D}, \\ M_{22} &= -\frac{B^2}{D} + 2G + \lambda, \\ M_{23} &= M_{32} = -\frac{CB}{D}, \\ M_{33} &= -\frac{C^2}{D} + G, \\ A &= G \cos 2\theta + (G + \lambda) \sin \varphi', \\ B &= G \cos 2\theta - (G + \lambda) \sin \varphi', \\ C &= G \sin 2\theta, \\ D &= G + (G + \lambda) \sin^2 \varphi', \end{aligned} \quad (2.17)$$

where G is shear modulus and λ is Lamé's constants and calculated by using

$$\lambda = \frac{2G\mu}{1 - 2\mu}. \quad (2.18)$$

For $d\sigma_y = 0$, $d\tau_{xy} = 0$, and $d\sigma'_y = -du$, the displacement of slop can be calculated by using

$$\begin{aligned} d\varepsilon_y &= -\frac{M_{33}}{M_{32}}d\gamma_{xy}, \\ d\gamma_{xy} &= -\frac{du \sin 2\theta [\cos 2\theta - (1 + \lambda/G) \sin \varphi']}{G (\cos 2\theta + \sin \varphi')^2 (1 + \lambda/G)}, \\ d\sigma'_x &= \left(-M_{12} \frac{M_{33}}{M_{32}} + M_{13} \right) d\gamma_{xy}. \end{aligned} \quad (2.19)$$

The slope angle, depth of impervious surface, depth initial groundwater level, weight density of groundwater, the sliding soil mass weight density, coefficient of lateral earth pressure, effective angle of internal friction, saturated hydraulic conductivity, and shear modulus were set up to $\alpha = 25^\circ$, $d = 4\text{ m}$, $d_w = 4\text{ m}$, $\gamma_w = 9.8\text{ kN/m}^3$, $\gamma_{\text{sat}} = 21.0\text{ kN/m}^3$, $k_0 = 0.7$, $\varphi' = 30^\circ$, $D_0 \approx 0.001\text{ m}^2/\text{s}$, and $G = 2\text{ MPa}$. Also, the coefficient of permeability is expressed as 10^{-4} m/s and the upper limit of light rain, moderate rain, heavy rain, heavy rainstorm, severe rainstorm is $1/K = 0.012, 0.025, 0.05, 0.1, \text{ and } 0.3$, respectively.

Figure 3 shows the effect of rainfall in slope displacement and the factor of slope safety (F_s) under different rainfall amount. There is an approximate one-to-one correspondence between displacement value and F_s with various grades of rainfall; F_s is found to decrease markedly with increasing displacement, and the influence of rainfall on the relationship between F_s and displacement becomes slight (Figure 3(a)). It needs about 124 hours, 53 hours, 26 hours, 14 hours, and 5 hours to light rain, moderate rain, heavy rain, heavy rainstorm and severe rainstorm to make 10 mm displacement of slop (Figure 3(b)); so rainfall has a huge impact on the displacement value.

During the calculation of the F_s , some parameters of model such as physical and mechanical parameters (effective angle of internal friction φ' and effective cohesion of soil c') are uncertain, and they are closely related to test method, soil specimens, reduction factor, and so on. So it is difficult to forecast the displacement of highway slope by this way. Because displacement has an approximate one-to-one correspondence to F_s , displacement can be used to forecast the failure of highway slop. And rainfall has important influence on the slop displacement; forecasting the displacement of cohesionless soil highway slope in the future with monitoring data based on artificial neural networks (ANNs) should consider the rainfall.

3. Methods of Study Using Modular Neural Network

The most feature of ANNs is that they are based on a self-organizing structure that resembles the biological neural system of mammalian brains, and ANNs give an efficient method to solve problems that cannot be exact solved by statistical methods for their theoretical limitations. Thus, displacement forecasting models based on ANNs ignore influence of rainfall, and the values might be unobjective and unconvincing.

MNN uses divide-and-conquer technique, in which a problem is divided into a set of subproblems according to the inherent class relations among training data, then gives a network module to learn each of the subproblems, recombine the individual modules into a solution to the original problem at last. Before learning displacement monitoring data,

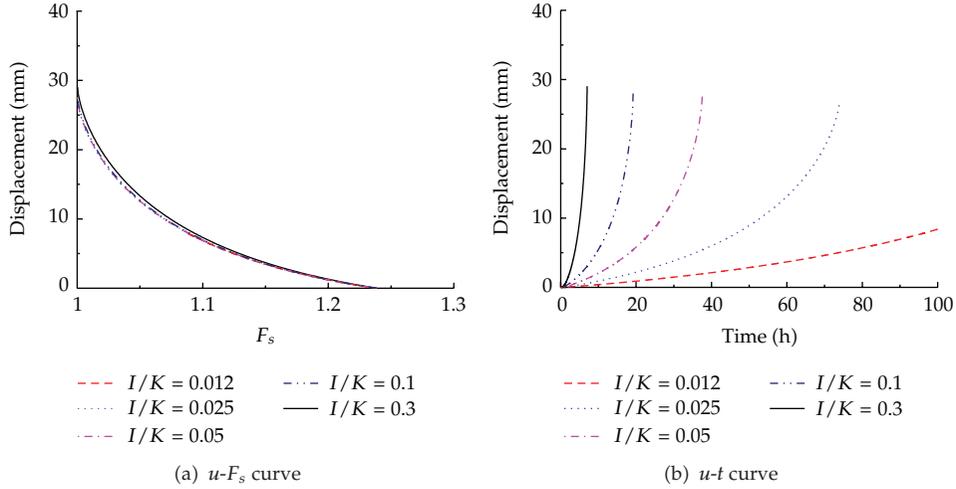


Figure 3: The effect of rainfall on displacement and F_s .

sample set of slope displacement monitoring data should be correctly classified based on the rainfall grade, after that construct the subnetwork to make the subsample data can be precisely trained, and finally recombine the subnetwork outputs into one. The total output of MNN is given by

$$y = \sum_{i=1}^R w_i y_i, \quad (3.1)$$

where y is the network output, y_i is the output of the i th subnetwork (denoted by NET_i) and w_i is the normalized output vector elements of the subnetwork, and R is the number of local networks.

Figure 4 shows the network architecture of the prediction system, which consists of three layers input layer, hidden layer, and output layer, and all of them are completely connected to form a hierarchical network. Subnetworks $\{NET_1, NET_2, \dots, NET_R\}$ constitute the network, state of the switches $\{E_1, E_2, \dots, E_R\}$ is identified by rainfall, and R is the number of local networks and switches. The sample set of slope displacement monitoring data is defined as $\mathbf{S} = \{s_1, s_2, \dots, s_n\}$, $\mathbf{Z} = \{x_1, x_2, \dots, x_n\}$ is the sample set of rainfall, and n is the number of data. \mathbf{S} and \mathbf{Z} have an approximate one-to-one correspondence.

The displacement-time forecasting model is established as (3.2), which allows to forecast the $(t + m)$ th displacement of slope in the future:

$$s_{t+m} = f(s_{t-1}, s_{t-2}, \dots, s_{t-q}), \quad (3.2)$$

where q is delay time step.

At first, construct eigenvector for rainfall $x = [LR, MR, HR, TR, SR]^T$, so that LR is light rain, MR is moderate rain, HR is heavy rain, TR is terrible rainstorm, and SR is severe

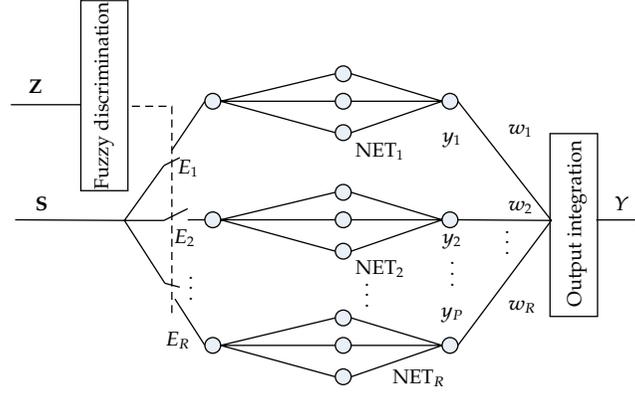


Figure 4: Modular neural network model.

rainstorm. Consider fuzzy classification of rainfall grade; the displacement monitoring data of sample set \mathbf{S} are classified according to calculation of eigenvector x using

$$V_i = \frac{\sum_{j=1}^N (u_{ij})^a x_j}{\sum_{j=1}^N (u_{ij})^a}, \quad (3.3)$$

$$v_{ij} = \frac{1}{\sum_{k=1}^R (\|x_j - V_i\| / \|x_j - V_d\|)^{2/(b-1)'}}$$

where R is the classification number of sample set, $\mathbf{S} = \{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_R\}$, \mathbf{V}_i is clustering center of \mathbf{S}_i , and $\mathbf{v} = (v_{ij})$ is the membership matrix of sample to fuzzy subset. In (3.3), $a = 1$ and $b = 2$ [17]. Some data may simultaneously belong to sample subsets \mathbf{S}_i and \mathbf{S}_j . After that, the back propagation algorithm is used to train the gating and local network NET_i with sample subset \mathbf{S}_i , and set lower bound on the error $\varepsilon = 10^{-8}$ and the maximum iteration number is 2000 [5]. Final results of the MNN can be calculated by (3.1).

To the new test data $s_f \notin \mathbf{S}$, the distance measure of x_f and $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_R]$ is chosen as criterion to decide which local network it belongs to, calculated using (3.4) and (3.5):

$$d_i = \frac{\|x_f - V_i\|}{\bar{d}_i}, \quad (3.4)$$

in which,

$$\bar{d}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \|x_j - V_i\|, \quad (3.5)$$

where d_i is the relative distance measure of x_f and NET_i , \bar{d}_i is the average distance measure of i th sample subset, and n_i is the number of data for i th sample subset. Displacement monitoring data s_f is trained only by local networks in which d_i is the smallest one based on

the traditional MNN it makes the calculation accuracy of boundary samples lower to others. To solve this problem, more than one local network are chosen to training s_f with fuzzy decision, in order to significantly enhance the accuracy of MNN model. First, the relative distance measure of x_f among all local network normalization can be calculated using

$$u_i = \frac{d_i}{\sum_{i=1}^R d_i}. \quad (3.6)$$

In (3.6), $\sum u_i = 1$ and $u_i \in [0, 1]$.

Defining fuzzy set of relative distance measure $\mathbf{A} = \{\text{very small (denoted by } vs), \text{small (denoted by } s), \text{middling (denoted by } m), \text{large (} l)\}$, the values of u_i reflect the distance measure between x_f and NET_i . If $u_i \in \{vs\}$, then local networks NET_i should be used to train s_f for the distance of x_f and NET_i is short. If $u_i \in \{l\}$, then local networks NET_i cannot be used to train s_f for the distance of x_f and NET_i is long.

The calculation of local networks selection involves the following steps.

Step 1. Calculate u_i between every new test data $s_f \notin \mathbf{S}$ and its corresponding local networks using (3.4), (3.5), and (3.6).

Step 2. Calculate membership degree of u_i to every fuzzy subset \mathbf{A} with the membership degree curves as Figure 5 shows.

Step 3. Make a choice which fuzzy subset u_i belongs to by the highest membership degree in the fuzzy command, and local networks which belong to the same set are chosen to integrate from vs to 1. If the local network in which $u_i \in \{vs\}$ actually exist, then it should be used to train the data s_f , otherwise, select local network in which $u_i \in \{s\}$, until selecting local network in which $u_i \in \{m\}$.

Due to the symmetry of the Gaussian distribution, the center value of membership degree decreases 50% at each time by adopting Gauss function for fuzzy membership functions. The center of fuzzy subset to membership degree $\{vs, s, m, l\}$ is $0.25/P, 0.5/P, 1/P, 2/P$ and the value of membership degree is 1 when $u_i < 0.25/P$ or $u_i > 2/P$. For $\sum u_i = 1$, all u_i cannot meet $u_i > 1/P$ at the same time; meanwhile, the local network corresponding to $u_i \in \{vs, s, m\}$ would not be the empty set. The local network corresponding to $u_i \in \{l\}$ should not be chosen for selecting sequence of local network is from vs to 1 in Step 3.

The weight of local network that is selected is given by

$$w_{fi} = \frac{1/d_{fi}}{\sum_{i=1}^L 1/d_{fi}}, \quad (3.7)$$

where w_{fi} is the weight of local network that is selected, d_{fi} is the distance measure of i th local network, and L is the number of local networks that has been selected.

4. Simulation Results

Hence, in order to illustrate the effectiveness of the model validity, one of Beijing city's highway slopes is chosen for test. Geological investigations in the collapse area in this

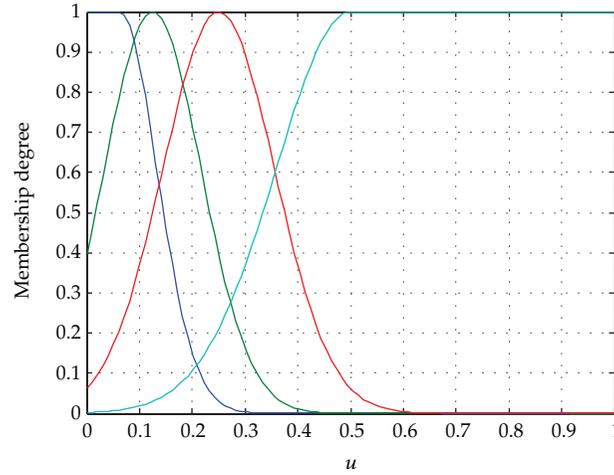


Figure 5: Membership degree curves of u_i .



Figure 6: Location of slope and brief geological descriptions.

research are based on site investigation and the evaluation of remote survey images; the location and brief geological descriptions of slope are shown as Figure 6.

Time series analysis was used to set up displacement-time forecasting model for both MNN model and ANN model [18]. And then forecast cohesionless soil highway slope displacement in the future based on the same sample data.

Setting model parameters $K = 5$ for the classification number of sample set, the number of hidden node in multilayer feedforward neural network that is used to train local networks is 4, and delay time step q which is calculated with autocorrelation method is 6 [19].

Figure 7 shows the relative error for the group of displacement monitoring data that is in the same rainfall grade, namely, ignoring influence of rainfall through selecting the displacement monitoring data in one rainfall grade. And their respective margins of error are almost the same as each other because both of the methods use back propagation algorithm to training.

Displacement data records the state of this cohesionless soil highway slope from the period May 22 to October 31, 2011. Due to the fact that rainfall series is necessary for the

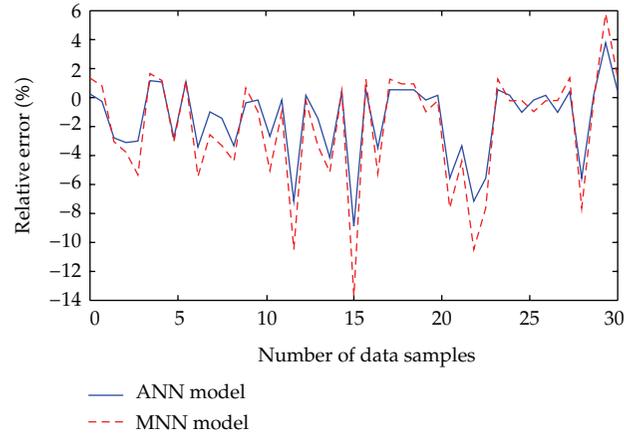


Figure 7: Relative error curve of models.

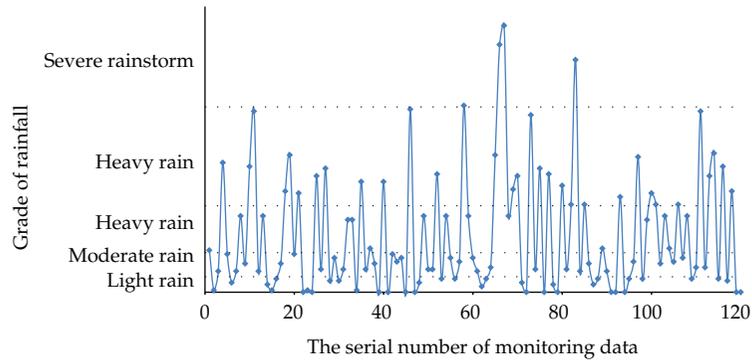


Figure 8: Rainfall distribution map.

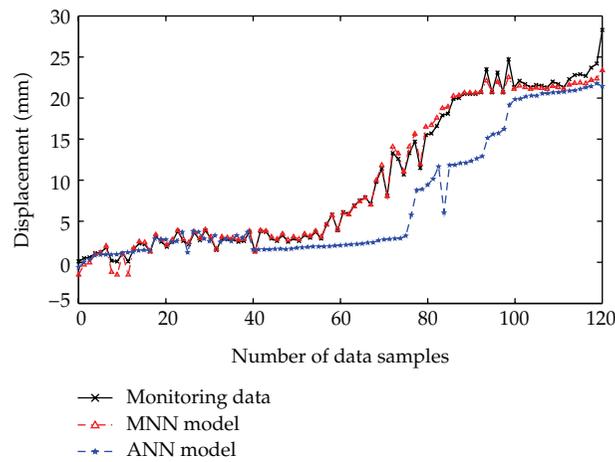


Figure 9: Displacement forecast curves.

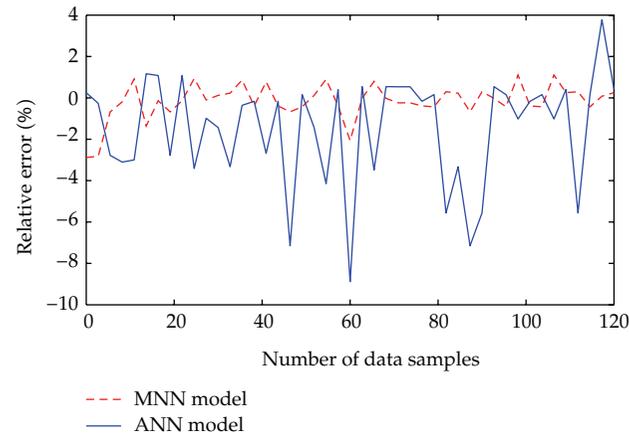


Figure 10: Relative error curve of models.

failure of slope, the meteorological station was built at the slope. Figure 8 shows the grade of rainfall distribution map at precipitation station of highway construction area during the whole observation period.

Figure 9 shows three curves of forecast data based on two methods and the real data during the whole observation period. The forecast data consider the influence of rainfall.

Figure 10 shows their relative error curves of displacement forecast data.

Note that in Figures 9 and 10, MNN offers a highly satisfactory method to forecast the slope displacement, which is better than the ANN model. MNN model has a moderate relative error, but ANN model shows even bigger fluctuations especially after intense rainfall.

The value of mean square error and average relative error of these models under different conditions is just as shown Table 1.

According to the above-mentioned analyses, we can get the results that (1) Whichever algorithm of artificial neural networks is chosen and both models have accurate forecast data of the slope displacement. (2) When considering the influence of rainfall, the fitting situation of MNN model is better than ANN model, mean square error of MNN model is only 7.73% of ANN model's, and average relative error of MNN model is 27.65% of ANN model's. (3) When considering the influence of rainfall, both models' errors are sharply decreased, and MNN model performs better than ANN model.

5. Conclusion

Displacement monitoring data of highway slop is classified by the MNN based on the same type of data with the same character, which fits the monitoring needed. Also, the method reduces the complexity of the modular neural network and its generalization capabilities could remarkably be enhanced.

When considering the influence of rainfall, the precision of MNN model is much higher than ANN model. The MNN model has a good applicability for the self-learning ability and data-driven ability than ANN model, and it can enhance the applicability of model by retraining the networks with new data at regular intervals.

Table 1: Accuracy of models.

Model	Ignore influence of rainfall		Consider the influence of rainfall	
	Mean square error/mm	Average relative error	Mean square error/mm	Average relative error
MNN model	7.927×10^{-2}	8.25%	2.737×10^{-4}	2.06%
ANN model	8.724×10^{-2}	6.17%	3.539×10^{-3}	7.45%

Acknowledgments

This research was supported by China's Eleventh Five-Year National Science and Technology Support Project (no. 2009BAG13A02) and Science and Technology Project of MOT (no. 2012364223300).

References

- [1] G. B. Crosta and F. Agliardi, "Failure forecast for large rock slides by surface displacement measurements," *Canadian Geotechnical Journal*, vol. 40, no. 1, pp. 176–191, 2003.
- [2] S. H. Ni, P. C. Lu, and C. H. Juang, "A fuzzy neural network approach to evaluation of slope failure potential," *Microcomputers in Civil Engineering*, vol. 11, no. 1, pp. 59–66, 1996.
- [3] S. Lee, J. H. Ryu, K. Min, and J. S. Won, "Landslide susceptibility analysis using GIS and artificial neural network," *Earth Surface Processes and Landforms*, vol. 28, no. 12, pp. 1361–1376, 2003.
- [4] K. M. Neaupane and S. H. Achet, "Use of backpropagation neural network for landslide monitoring: A case study in the higher Himalaya," *Engineering Geology*, vol. 74, no. 3-4, pp. 213–226, 2004.
- [5] H. B. Wang, W. Y. Xu, and R. C. Xu, "Slope stability evaluation using Back Propagation Neural Networks," *Engineering Geology*, vol. 80, no. 3-4, pp. 302–315, 2005.
- [6] L. Ermini, F. Catani, and N. Casagli, "Artificial Neural Networks applied to landslide susceptibility assessment," *Geomorphology*, vol. 66, no. 1–4, pp. 327–343, 2005.
- [7] H. Guo, W. Wang, W. Guo et al., "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [8] J. B. Zhao, L. Li, and Q. Gao, "Research and application of the grey theory to slope deformation prediction," *Chinese Journal of Rock Mechanics and Engineering*, vol. 24, no. 2, pp. 5799–5802, 2005.
- [9] D. Wang, M. Zhou, and Y. Li, "Displacement analysis of slope based on wavelet transformation and grey theory," in *Proceedings of the International Conference on Electric Technology and Civil Engineering (ICETCE '11)*, pp. 4929–4932, Lushan, China, April 2011.
- [10] W. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, China Communications Press, Beijing, China, 2012.
- [11] H. Zhi-quan and F. U. Wen-bin, "The whole-region method of chaotic time series for slope deformation prediction," *Journal of North China Institute of Water Conservancy and Hydroelectric Power*, vol. 30, no. 3, pp. 70–73, 2009.
- [12] M. B. Jaksa, *The influence of spatial variability on the geotechnical design properties of a stiff, overconsolidated clay [Ph.D. thesis]*, The University of Adelaide, Adelaide, Australia, 1995.
- [13] T. Glade, M. Crozier, and P. Smith, "Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical "Antecedent Daily Rainfall Model"," *Pure and Applied Geophysics*, vol. 157, no. 6-8, pp. 1059–1079, 2000.
- [14] J. M. Gasmo, H. Rahardjo, and E. C. Leong, "Infiltration effects on stability of a residual soil slope," *Computers and Geotechnics*, vol. 26, no. 2, pp. 145–165, 2000.
- [15] C. Chien-Yuan, C. Tien-Chien, Y. Fan-Chieh, and L. Sheng-Chi, "Analysis of time-varying rainfall infiltration induced landslide," *Environmental Geology*, vol. 48, no. 4-5, pp. 466–479, 2005.
- [16] R. M. Iverson, "Landslide triggering by rain infiltration," *Water Resources Research*, vol. 36, no. 7, pp. 1897–1910, 2000.
- [17] M. G. C. A. Cimino, W. Pedrycz, B. Lazzarini, and F. Marcelloni, "Using multilayer perceptrons as receptive fields in the design of neural networks," *Neurocomputing*, vol. 72, no. 10–12, pp. 2536–2548, 2009.

- [18] X. Liu, X. H. Zeng, and C. Y. Liu, "Research on artificial neural network-time series analysis of slope nonlinear displacement," *Chinese Journal of Rock Mechanics and Engineering*, vol. 24, no. 19, pp. 3499–3504, 2005.
- [19] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica D*, vol. 73, no. 1-2, pp. 82–98, 1994.

Research Article

A Fuzzy Optimization Model for High-Speed Railway Timetable Rescheduling

Li Wang,^{1,2} Yong Qin,¹ Jie Xu,¹ and Limin Jia¹

¹ State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

² School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Limin Jia, jialm@vip.sina.com

Received 19 September 2012; Revised 19 November 2012; Accepted 21 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 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 fuzzy optimization model based on improved symmetric tolerance approach is introduced, which allows for rescheduling high-speed railway timetable under unexpected interferences. The model nests different parameters of the soft constraints with uncertainty margin to describe their importance to the optimization purpose and treats the objective in the same manner. Thus a new optimal instrument is expected to achieve a new timetable subject to little slack of constraints. The section between Nanjing and Shanghai, which is the busiest, of Beijing-Shanghai high-speed rail line in China is used as the simulated measurement. The fuzzy optimization model provides an accurate approximation on train running time and headway time, and hence the results suggest that the number of seriously impacted trains and total delay time can be reduced significantly subject to little cost and risk.

1. Introduction

The infrastructures of high-speed railway have been extensively developed in China for the past several years. The network topology structure and operation mode of the railway are changing profoundly. The target is to cover its major economic areas with a high-speed railway network, which consists of four horizontal and four vertical lines [1], in the following several years. The network scale is much larger than any existing ones in the world. However, passengers have to face huge challenging variables generated by unexpectedness, for example, weather, equipment failure. On June 20, 2010, for example, the southern part of the existing Beijing-Shanghai line experienced speed restriction for a heavy rain storm. As announced, 18 trains departing from Shanghai were canceled, thereby more than 20,000 passengers being affected. Also, the rest in-service trains were more or less delayed. Considering the high speed and frequency of bullet train, the impact of train delay and

speed restriction would be more serious than that 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, train timetable rescheduling is a focal point to improve the operation mode with the characteristics of high train speed, high train frequency, and mixed train speed (HHM), which is of great theoretical and practical significance for safe and efficient operation of China high speed railway network.

Train rescheduling in China is mostly manually developed by the operator with the support of computer in the existing line or the high speed railway, which depends on individual's experience to dominate single line or some sections in one line with fixed three-hour period. In future, China will establish six comprehensive operation centers of the railway network, each of which will dominate several lines covering one thousand kilometers overall. The scale of rescheduling object is far larger, and the relationship between lines is far more complex than the existing situation, so the inefficient manual rescheduling, which seriously affects the use of the ability of high-speed rail and brings risk to railway safety, cannot support the advanced operation command mode.

The research on the train operation automatic adjustment has been a controversial topic for many years in literature. Variable academic algorithms for optimal rescheduling have been put forward. Some traditional optimization model, such as integer programming (branch and bound method [2–5]) and linear programming [6–9], are usually used for the single transportation organization mode (only high-speed or medium-speed train running). In China, there are usually mixed high-speed and medium-speed trains running in one line, so there are great differences on the variable size, constraints, and target in the actual situation of the Chinese high-speed railway, and these methods cannot be directly applied to the train rescheduling problems in China.

Since the railway network has been extensively developed, and the operation mode is becoming more complex in recent years, many heuristic methods, such as DEDS-based simulation method [10], expert system [11, 12], tabu search [13–15], and some computational intelligence methods, such as genetic algorithm [16–18], particle swarm optimization algorithm [19–21], and other composite algorithms [22–26], were used to solve the large-scale combinatorial optimization problem by many scholars. To some extent, these mathematical models and optimization algorithms have given feasible solutions for train operation automatic adjustment problem. However, the decision variables, parameters, and constraints are far more complex, as they affect on the goals for high-speed railway timetable rescheduling optimization. For example, typical uncertain variables, such as the running time at a section and the dwell time at a station of the train, are easily affected by railway and equipment status, driving behavior of drivers, and environment condition [27, 28]. As a result, these methods are limited on applying to the actual operation environment of high-speed railway under the significant influence of uncertainty, and hence they cannot be considered as an appreciative approach to either achieve the purpose of automatic rescheduling or play the role of decision support in practical applications. Therefore, the uncertainty character of high-speed train automatic adjustment model should be considered in substance.

Uncertainty optimization theory is to solve the optimization decision problem with all kinds of uncertainties, which involves stochastic optimization, fuzzy optimization, rough set, and so on. Nowadays, some papers studied uncertainty in train scheduling problem. Jia and Zhang [29] proposed a distributed intelligent railway traffic control system based on fuzzy decision making. Yang et al. [30] investigates a passenger train timetable problem

with fuzzy passenger demand on a single line railway. Since the number of passengers getting on/off the train at each station is assumed to be a fuzzy variable, the total passengers time is also a fuzzy variable. An expected value goal-programming model is constructed to minimize the total passenger time and the total delay. A branch-and-bound algorithm based on the fuzzy simulation is designed in order to obtain an optimal solution. Then they extend the uncertainty in [31] by considering stochastic and fuzzy parameters synchronously to solve the railway freight transportation planning problem. Based on the chance measure and critical values of the random fuzzy variable, three chance-constrained programming models are constructed for the problem with respect to different criteria. A fuzzy periodic job shop scheduling model is introduced to address the framework of the periodic robust train scheduling problem in [32]. Fuzzy approach is used to reach a tradeoff among the total train delays, the robustness of schedules, and the time interval between departures of trains from the same origins in this paper. Cucala et al. [33] proposes a fuzzy linear programming model to minimize energy consumption with uncertain delays and drivers behavioral response. The method is applied to a real Spanish high speed line to optimize the operation, and comparing to the current commercial service evaluates the potential energy savings. Wang et al. [34, 35] and Guo et al. [36] analyzed the driver's safety approaching behaviour and pedestrian safety crossing behaviour in the urban traffic environment. The traffic participants revealed different behavioral decisions with various personal characteristics and can be described as fuzzy parameters. Overall, none of the previous papers analyzed the uncertainty of all factors in train rescheduling, although they considered some constraints as fuzzy member, such as passenger time and delay time. Furthermore, the fuzzy constraints are defined as triangular or trapezoidal fuzzy numbers, which are very difficult to determine the distribution.

In this paper, we attempt to achieve optimal timetable rescheduling under the uncertainties, for example, constraints and/or unexpected parameters, by means of proposing the fuzzy optimization model as discussed in the above. In the rest of this paper, typical rescheduling model will be discussed in the section of Timetable Rescheduling Problem. In the following section, we will describe in details of the fuzzy optimization model based on improved tolerance approach to timetable rescheduling, including the fuzzy membership functions of the original objective and soft constraints. A case study on the busiest section of Beijing-Shanghai high speed line will be illustrated in the section of Case Study. The final section concludes the results of the paper and suggests for further research.

2. Timetable Rescheduling Problem

The aim of train rescheduling is to get a new timetable that adjusts the train movements to be consistent with the planned schedule as much as possible under some interference [37]. The following model focuses on minimizing the total delay as well as the number of seriously impacted trains.

2.1. Input Data

Take a rail line with n trains and m stations for example. The numerical inputs are described as follows:

S_j : the station j according to original timetable.

$x_{i,j}^*$: the departure time of train i at station j according to original timetable.

$y_{i,j}^*$: the arrival time of train i at station j according to original timetable.

$a_{i,j}$: the minimum running time of train i on section $[j, j + 1]$.

τ_j^f : the time interval between two adjacent trains at station j .

τ_j^d : the headway of section $[j, j + 1]$.

τ_j^{tr} : the track number of station j .

$T_{i,j}$: the minimum dwell time of the train i at station j .

$h_{i,j}$: if train i stop at station j according to original timetable, then $h_{i,j} = 1$; otherwise, $h_{i,j} = 0$.

c_i^{delay} : the cost per time unit delay for train i .

T_i^D : delay tolerance for train i .

θ : weight for objective.

M : a very big integer, for example, 100000.

2.2. Decision Variables

The decision variables are described as follows:

$x_{i,j}$: the new departure time of train i at station j after adjustment.

$y_{i,j}$: the new arrival time of train i at station j after adjustment.

$d_{i,j}$: delay of train i at station j , which is defined as the difference between the arrival time after adjustment and the planned arrival time in the original timetable.

b_i : if train reaches its final considered stop with a delay larger than w_i , then $b_i = 1$; otherwise, $b_i = 0$.

$\eta_{i,j,l}$: if train i use track l at station j , then $\eta_{i,j,l} = 1$; otherwise, $\eta_{i,j,l} = 0$.

$\alpha_{i,j,i',j'}$: if the departure time of train i at station j is earlier than the departure time of train i' at station j' to original timetable, then $\alpha_{i,j,i',j'} = 1$; otherwise, $\alpha_{i,j,i',j'} = 0$.

$\beta_{i,j,i',j'}$: if the departure time of train i at station j is changed to occur after the departure time of train i' at station j' , then $\beta_{i,j,i',j'} = 1$; otherwise, $\beta_{i,j,i',j'} = 0$.

2.3. Objective Functions

(i) To minimize the delay cost:

$$\text{Minimize } \sum_{i=1}^n \left(c_i^{\text{delay}} \cdot \sum_{j=1}^m d_{i,j} \right). \quad (2.1)$$

(ii) To minimize the number of seriously impacted trains:

$$\text{Minimize } \sum_{i=1}^n b_i. \quad (2.2)$$

We set the final objective function as:

$$S = \theta \sum_{i=1}^n \left(c_i^{\text{delay}} \cdot \sum_{j=1}^m d_{i,j} \right) + (1 - \theta) \sum_{i=1}^n b_i. \quad (2.3)$$

2.4. Constraints

(i) Section running time restrictions:

$$y_{i,j+1} \geq x_{i,j} + a_{ij}. \quad (2.4)$$

The real departure time cannot be earlier than the original departure time:

$$\begin{aligned} x_{i,j} &\geq x_{i,j}^*, & h_{i,j} &= 1, \\ x_{i,j} - x_{i,j}^* &= d_{i,j}. \end{aligned} \quad (2.5)$$

(ii) Station dwell time restrictions:

$$x_{i,j} - y_{i,j} \leq T_{i,j}. \quad (2.6)$$

(iii) Track restrictions:

$$\sum_{l=1}^{\tau_j^{\text{tr}}} \eta_{i,j,l} = 1. \quad (2.7)$$

(iv) Station headway restrictions.

For each station, if two trains use the same track, at least one of α and β is forced to be 1:

$$\eta_{i,j,l} + \eta_{i',j,l} - 1 < \alpha_{i,j,i',j'} + \beta_{i,j,i',j'}, \quad (2.8)$$

$$y_{i',j} - y_{i,j} \leq \tau_j^f \alpha_{i,j,i',j'} - M(1 - \alpha_{i,j,i',j'}), \quad (2.9)$$

$$y_{i,j} - y_{i',j} \leq \tau_j^f \beta_{i,j,i',j'} - M(1 - \beta_{i,j,i',j'}). \quad (2.10)$$

(v) Section headway restrictions.

For each section, at least one of α and β is forced to be 1 because there is only one track:

$$\alpha_{i,j,i',j'} + \beta_{i,j,i',j'} = 1, \quad (2.11)$$

$$x_{v,j} - x_{i,j} \leq \tau_j^d \alpha_{i,j,i',j'} - M(1 - \alpha_{i,j,i',j'}), \quad (2.12)$$

$$x_{i,j} - x_{v,j} \leq \tau_j^d \beta_{i,j,i',j'} - M(1 - \beta_{i,j,i',j'}). \quad (2.13)$$

(vi) Auxiliary restrictions:

$$d_{\text{lost}(i)} - T_i^D \leq Mb_i. \quad (2.14)$$

In practice, some constraints of the model are not strictly satisfied due to the inexact operation time. Thus four of the above constraints need to be changed as below.

(i) *Section Running Time Restrictions*

$a_{i,j}$ is the minimum running time of train i in section $[j, j+1]$, including the pure running time in section and the additional time for train stop or departure at the station. The minimum running time is decided by the length of the section, the infrastructure characters of the section, and the train type. In the actual operation environment, the train running speed is not a constant value because of the various infrastructure characters of the railway, that is, bridges, tunnels, and culverts. Furthermore, some factors (i.e., railway equipment statuses, technological level of crew, and weather condition) also increase the uncertainty of running time. Thus it is very important to find a safe and reasonable average speed that is far below the limited speed and full utilization of railway capacity, which will greatly improve the optimization result especially when some trains are delayed due to some interference. Since the minimum running time of the train in the section usually changes within a certain range, there should be a tolerance for $a_{i,j}$. Thus (2.4) does not need to be strictly satisfied and can be changed as below:

$$g_1 = y_{i,j+1} - x_{i,j} \leq a_{i,j}. \quad (2.15)$$

(ii) *Station Dwell Time Restrictions*

$T_{i,j}$ is the minimum dwell time of the train i at station j , including the pure operation time at station, passengers on and off time, crew setup time, and some additional time, like waiting for other trains. The minimum dwell time is decided by station operation type, station level, and the train type. In the actual operation environment, railway equipment statuses, driving behavior of drivers, and environment condition may increase the uncertainty of running time. Thus it is also important to find a safe and reasonable average dwell time that can be full

utilization of station capacity. Similar to section running time, there also should be a tolerance for $T_{i,j}$, and the dwell time constraint can be changed as below:

$$g_2 = x_{i,j} - y_{i,j} - T_{i,j}. \quad (2.16)$$

(iii) Station Headway Restrictions

The headway time for each station is decided by the number of receiving-departure track, the operation time of turnout, and holding time of the track, which also faces uncertainty problems due to the factors of equipment status and human technological level. So there should be a tolerance for τ_j^f . Equations (2.9) and (2.10) do not need to be strictly satisfied and can be changed as below:

$$\begin{aligned} g_5 &= y_{i',j} - y_{i,j} - \tau_j^f \alpha_{i,j,i',j'} - M(1 - \alpha_{i,j,i',j'}), \\ g_6 &= y_{i,j} - y_{i',j} - \tau_j^f \beta_{i,j,i',j'} - M(1 - \beta_{i,j,i',j'}). \end{aligned} \quad (2.17)$$

(iv) Section Headway Restrictions

The headway time for each section is decided by the number and length of the block between two tracking trains and the speeds of the trains, which ranges from 2.4 minutes to 3 minutes according to Shi [38]. Similar to station headway constraint, (2.12) and (2.13) can be changed as below:

$$\begin{aligned} g_3 &= x_{i',j} - x_{i,j} - \tau_j^d \alpha_{i,j,i',j'} - M(1 - \alpha_{i,j,i',j'}), \\ g_4 &= x_{i,j} - x_{i',j} - \tau_j^d \beta_{i,j,i',j'} - M(1 - \beta_{i,j,i',j'}). \end{aligned} \quad (2.18)$$

Since the four constraints are changed as above, the model turns to be not representative in the sense of mathematical viewpoints. To construct a reasonable mathematical model under the uncertain environment, the tolerance approach based timetable rescheduling model will be introduced in the next section.

3. Fuzzy Optimization Model for Timetable Rescheduling

In the paper, we use the fuzzy optimization based improved tolerance approach to solve the uncertainty program, and some necessary backgrounds and notions of the approach are reviewed.

3.1. Improved Tolerance Approach

Tolerances are indicated in any technical process, that is, the admissible limit of variation around the object value and the deviations allowed from the specified parameters [39].

A general model of a fuzzy linear programming problem is presented by the following system [40]. $\tilde{A}_{i,j}$, \tilde{B}_i , and \tilde{C}_j ($i = 1 \cdots m; j = 1 \cdots n$) are fuzzy set in R . The symbol \oplus represents the extended addition. Each real number can be modeled as a fuzzy number.

$$\begin{aligned} & \text{Minimize } \tilde{C}_1x_1 \oplus \tilde{C}_2x_2 \oplus \cdots \oplus \tilde{C}_nx_n, \\ & \text{Subject to } \tilde{A}_{i,1}x_1 \oplus \tilde{A}_{i,2}x_2 \oplus \cdots \oplus \tilde{A}_{i,n}x_n \quad \tilde{B}_i, \quad i = 1 \cdots m. \end{aligned} \quad (3.1)$$

Here we only discuss the special case (3.2) where the objective function is crisp, some constraints have the soft form, and the rest constraints are crisp.

$$\begin{aligned} & \text{Minimize } z(x) = c_1x_1 + c_2x_2 + \cdots + c_nx_n, \\ & g_i(x) = a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n \quad \tilde{B}_i, \quad i = 1 \cdots m_1, \\ & a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n \quad b_i, \quad i = m_1 + 1 \cdots m, \\ & x_1, \dots, x_n \quad 0. \end{aligned} \quad (3.2)$$

The m_1 soft constraints may be described more precisely by the fuzzy set with the support $[b_i - d_i, b_i]$, and d_i is the tolerance according to b_i . Moreover the membership function of $g_i(x)$ can be specified as (3.3), and its graphics is Figure 1(a). Then we can directly assigns a measure of the satisfaction of the i th constraint to the solution $X = (x_1, x_2, \dots, x_n)$.

$$\mu(g_i) = \begin{cases} 0, & g_i > b_i - d_i, \\ \frac{g_i - (b_i - d_i)}{d_i}, & b_i - d_i < g_i < b_i, \\ 1, & g_i < b_i. \end{cases} \quad i = 1 \cdots m_1, \quad (3.3)$$

Fuzzy constraints will inevitably lead to the fuzzy objective based on the ideology of symmetric model. Thus, for the objective function $z(x)$, there is a fuzzy set $\tilde{Z} = \{(z, \mu_z(z)) \mid z \in R\}$. Let $X_l = \{x \in R \mid a_{i,1}x_1 + \cdots + a_{i,n}x_n \leq b_i - d_i, \text{ for all } i = 1 \cdots m_1 \text{ and } a_{i,1}x_1 + \cdots + a_{i,n}x_n \leq b_i, \text{ for all } i = m_1 + 1 \cdots m\}$, $X_u = \{x \in R \mid a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n \leq b_i, \text{ for all } i = 1 \cdots m\}$, $\bar{z} = \text{Min}_{x \in X_u} z(x)$, and $\underline{z} = \text{Min}_{x \in X_l} z(x)$. So the membership function $\mu_z(z)$ of \tilde{Z} is given by (3.4), and its graphics is Figure 1(b).

$$\mu_z(z) = \begin{cases} 0, & z > \bar{z}, \\ 1 - \frac{z - \underline{z}}{\bar{z} - \underline{z}}, & \underline{z} < z < \bar{z}, \\ 1, & z < \underline{z}. \end{cases} \quad (3.4)$$

In order to determine a compromise solution, it is usually assumed that the total satisfaction of a decision maker may be described by $\lambda(x) = \min(\mu_z(x), \mu_1(x), \mu_{m_1}(x))$. Since not all the constraints are equally important to the solution in practice, we take λ_i to describe the priority of the i th soft constraint to the solution, and $\lambda(x)$ can be described by $\lambda(x) = \sum_1^{m_1} w_i \lambda_i(x)$. w_i is the weight of λ_i , which shows the importance of different soft

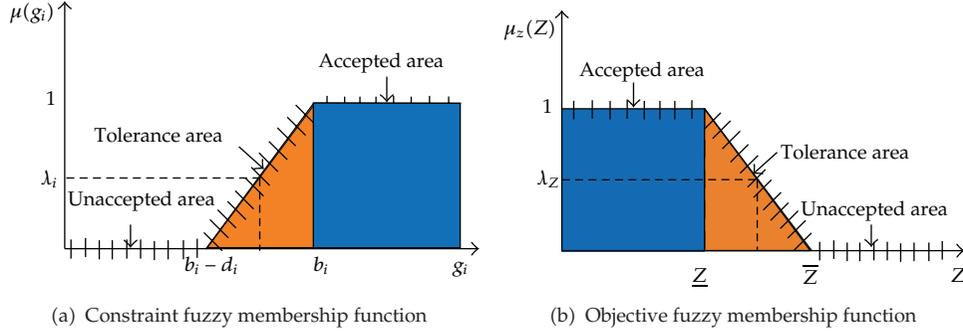


Figure 1: Fuzzy membership functions.

constraints to the optimization purpose. Generally, there are many complex constraints for different optimization systems in practice. It is very hard to handle all constraints during the system optimization, and therefore the systems always treat the constraints with different priorities. Moreover the decision makers also hold different views on the importance of various factors about the same issue. The timetable rescheduling problem, for example, will take different optimization objectives and constraint priorities in different emergencies. The dispatcher will focus on the reducing of train delay time and train headway time if few trains are disturbed by equipment failure. However, minimizing the number of seriously impacted trains and the train interval running time as much as possible are more effective measures to return to the normal states, when some accidents or natural disasters lead to speed restriction on lots of sections for a long time.

Since we treat the objective in the same manner as the soft constraints, this is also a symmetric model [41]. More information about symmetric fuzzy model can refer to [42]. The optimization program is clearly equivalent to (3.5). Using linear membership functions or piecewise linear, concave membership function, the system can easily be solved by well-known algorithms.

$$\begin{aligned}
 & \text{Maximize } \lambda, \\
 & \mu_z(x) \geq \lambda, \\
 & \mu_i(x) \geq \lambda_i, \quad i = 1 \cdots m_1, \\
 & a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n \leq b_i, \quad i = m_1 + 1 \cdots m, \\
 & x \in X, \quad \lambda_i \in [0, 1].
 \end{aligned} \tag{3.5}$$

Figure 2 shows the principle of the tolerance based symmetric fuzzy model. When the constraint relaxes to b^1 , the objective value minimized to Z^1 (a significant improvement). However, when the constraint relaxes to b^2 , the objective value improves little ($Z^1 - Z^2$) compare to the slack of constraint ($b^2 - b^1$). The tolerance-based symmetric fuzzy model is to get the maximum value of the fuzzy member λ , such that the system can obtains a great improvement of the objective on the conditions of less relaxation of the constraints. This means, for the train adjustment problem, we can get a new timetable to eliminate interference as much as possible with little slack of train rescheduling constraints.

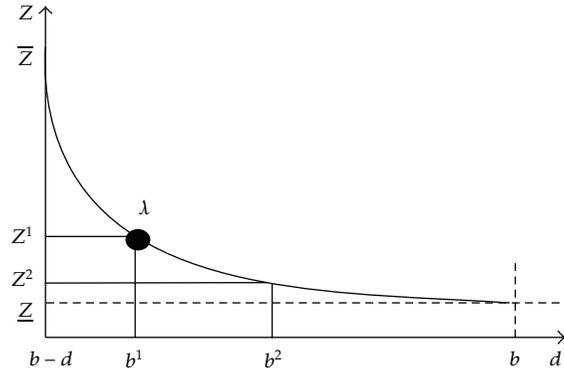


Figure 2: The principle of the fuzzy model.

3.2. A New Timetable Rescheduling Model

In the rescheduling model mentioned above, four additional inputs are involved to describe the tolerances of minimum running time, dwell time, and headway time for section and station, respectively. An additional decision variable λ is used to describe the fuzzy number for the objective and constraints. All the additional parameters are listed as below:

$c_{a_{i,j}}$: tolerance of $a_{i,j}$.

$c_{T_{i,j}}$: tolerance of $T_{i,j}$.

$c_{\tau_j^d}$: tolerance of τ_j^d .

$c_{\tau_j^f}$: tolerance of τ_j^f .

λ_i : decision variable for the fuzzy number.

s^0 : objective value of the original model based on original constraints.

s^* : objective value of the original model based on relaxed constraints.

Fuzzy membership functions of objective and soft constraints are given by μ_0 to μ_6 (3.6); their graphics are similar to Figure 1. The timetable rescheduling model is changed as (3.7). Some principles for the relaxation extent of the constraints are summarized as follows. (1) Collect statistics of train running states in all kinds of emergencies in reality. (2) Consult with the train drivers and dispatchers, who have a wealth of experience about train driving

and operation. (3) Refer to the train safety operation specifications under different emergency conditions.

$$\begin{aligned}
 \mu_0 &= \begin{cases} 1, & S < s^*, \\ 1 - \frac{S - s^*}{s_0 - s^*}, & s^* \leq S \leq s_0, \\ 0, & S > s_0, \end{cases} & \mu_1 &= \begin{cases} 1, & g_1 > a_{i,j}, \\ 1 - \frac{a_{i,j} - g_1}{c_{a_{i,j}}}, & a_{i,j} - c_{a_{i,j}} \leq g_1 \leq a_{i,j}, \\ 0, & g_1 < a_{i,j} - c_{a_{i,j}}, \end{cases} \\
 \mu_2 &= \begin{cases} 1, & g_2 > T_{i,j}, \\ 1 - \frac{T_{i,j} - g_2}{c_{T_{i,j}}}, & T_{i,j} - c_{T_{i,j}} \leq g_2 \leq T_{i,j}, \\ 0, & g_2 < T_{i,j} - c_{T_{i,j}}, \end{cases} & \mu_3 &= \begin{cases} 1, & g_3 > \tau_j^d, \\ 1 - \frac{\tau_j^d - g_3}{c_{\tau_j^d}}, & T_{i,j} - c_{\tau_j^d} \leq g_3 \leq \tau_j^d, \\ 0, & g_3 < T_{i,j} - c_{\tau_j^d}, \end{cases} \\
 \mu_4 &= \begin{cases} 1, & g_4 > \tau_j^d, \\ 1 - \frac{\tau_j^d - g_4}{c_{\tau_j^d}}, & T_{i,j} - c_{\tau_j^d} \leq g_4 \leq \tau_j^d, \\ 0, & g_4 < T_{i,j} - c_{\tau_j^d}, \end{cases} & \mu_5 &= \begin{cases} 1, & g_5 > \tau_j^f, \\ 1 - \frac{\tau_j^f - g_5}{c_{\tau_j^f}}, & T_{i,j} - c_{\tau_j^f} \leq g_5 \leq \tau_j^f, \\ 0, & g_5 < T_{i,j} - c_{\tau_j^f}, \end{cases} \\
 \mu_6 &= \begin{cases} 1, & g_6 > \tau_j^f, \\ 1 - \frac{\tau_j^f - g_6}{c_{\tau_j^f}}, & T_{i,j} - c_{\tau_j^f} \leq g_6 \leq \tau_j^f, \\ 0, & g_6 < T_{i,j} - c_{\tau_j^f}, \end{cases}
 \end{aligned} \tag{3.6}$$

Maximize λ ,

$$\theta \sum_{i=1}^n \left(c_i^{\text{delay}} \cdot \sum_{j=1}^m d_{i,j} \right) + (1 - \theta) \sum_{i=1}^n b_i \quad s^* + (1 - \lambda)(s^0 - s^*),$$

$$\lambda = \sum_1^4 w_i \lambda_i,$$

$$y_{i,j+1} - x_{i,j} \quad a_{i,j} + (\lambda_1 - 1)c_{a_{i,j}},$$

$$x_{i,j} \quad x_{i,j}^*, \quad h_{i,j} = 1,$$

$$x_{i,j} - x_{i,j}^* = d_{i,j},$$

$$x_{i,j} - y_{i,j} \quad T_{i,j} + (\lambda_2 - 1)c_{T_{i,j}},$$

$$\sum_{l=1}^{\tau_j^{\text{tr}}} \eta_{i,j,l} = 1,$$

$$\eta_{i,j,l} + \eta_{i',j,l} - 1 < \alpha_{i,j,i',j} + \beta_{i,j,i',j},$$

$$y_{i',j} - y_{i,j} \quad \left(\tau_j^f + (\lambda_3 - 1)c_{\tau_j^f} \right) \alpha_{i,j,i',j} - M(1 - \alpha_{i,j,i',j}),$$

$$y_{i',j} - y_{i,j} \quad \left(\tau_j^f + (\lambda_3 - 1)c_{\tau_j^f} \right) \beta_{i,j,i',j} - M(1 - \beta_{i,j,i',j}),$$

$$\begin{aligned}
& \alpha_{i,j,i',j} + \beta_{i,j,i',j} = 1, \\
& x_{i',j} - x_{i,j} - \left(\tau_j^d + (\lambda_4 - 1)c_{\tau_j^d} \right) \alpha_{i,j,i',j} - M(1 - \alpha_{i,j,i',j}), \\
& x_{i',j} - x_{i,j} - \left(\tau_j^d + (\lambda_4 - 1)c_{\tau_j^d} \right) \beta_{i,j,i',j} - M(1 - \beta_{i,j,i',j}), \\
& d_{\text{lost}(i)} - T_i^D - Mb_i.
\end{aligned} \tag{3.7}$$

Since M is a very integer, the headway time constraints can be changed as below:

$$\begin{aligned}
& y_{i',j} - y_{i,j} - \tau_j^f + (\lambda_3 - 1)c_{\tau_j^f} - M(1 - \alpha_{i,j,i',j}), \\
& y_{i',j} - y_{i,j} - \tau_j^f + (\lambda_3 - 1)c_{\tau_j^f} - M(1 - \beta_{i,j,i',j}), \\
& x_{i',j} - x_{i,j} - \tau_j^d + (\lambda_4 - 1)c_{\tau_j^d} - M(1 - \alpha_{i,j,i',j}), \\
& x_{i',j} - x_{i,j} - \tau_j^d + (\lambda_4 - 1)c_{\tau_j^d} - M(1 - \beta_{i,j,i',j}).
\end{aligned} \tag{3.8}$$

4. Case Study

This model is simulated on the busiest part of Beijing to Shanghai high speed line, between Nanjing (Ning for short) and Shanghai (Hu for short). In the rest of this paper, ‘‘Hu-Ning Section’’ is used to represent this part of the Beijing to Shanghai high speed line. There are seven stations on the Hu-Ning Section, thereby there are 6 sections whose lengths are 65110 m, 61050 m, 56400 m, 26810 m, 31350 m, and 43570 m. Its daily service starts at 6:30 am and ends at 11:30 pm. As currently planned, there are 52 trains (14 high speed trains and 38 medium speed trains) from Beijing to Shanghai line and 8 extra medium speed trains from Riverside line that go from Ning to Hu by the Hu-Ning Section. The trains from Beijing to Shanghai line are called self-line trains, while those from Riverside line are called cross-line trains. 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. Since Beijing to Shanghai high speed line is double track, we only consider the direction from Ning to Hu without loss of generality. The experimental procedure is divided into two stages for the full proof of the validity of the fuzzy model. Firstly, we do research in six aspects with different fuzzy constraints of the same weight, which proves the effectiveness of the tolerance-based fuzzy model in different trains operation conditions. Then a sensitivity analysis of the weighing factors is realized based on the previous operation circumstances. All the models are solved by Ilog Cplex 12.2.

4.1. Same Weight

4.1.1. Delay of One Train

In the simulation, we assume that the train G103 is late for 20 minutes in the section from Nanjing to Zhenjiang, and the trains can run at the normal speed in all the sections;

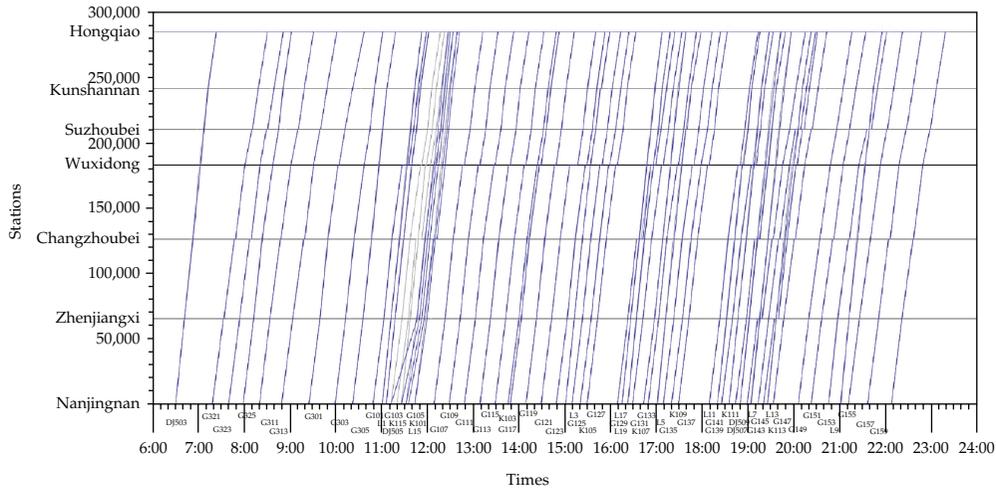


Figure 3: Timetable of one train delay.

the headway is set from 3 minutes to 2.5 minutes, and the minimum separation time on track possession in each station is set to 1 min.

First we use the original model to solve the problem and get the follow results. When the speed of high-speed train is set to 360 km/h, the speed of medium high-speed train is set to 320 km/h, the headway time is defined as 3 minutes, and the solution value is 368784; there are 3 trains late with the delay time between 10 minutes to 20 minutes and 1 train late with the delay time between 20 minutes to 30 minutes; and the total delay time is 47.5 minutes. When the speed of high-speed train is set to 380 km/h, the speed of medium high-speed train is set to 350 km/h, the headway time is defined as 2.5 minutes, and the solution value is 350449; there are one train late with the delay time between 0 minutes to 10 minutes and 2 trains late with the delay time between 10 minutes to 20 minutes; and the total delay time is 31.85 minutes. Then we use the above results as the inputs of the fuzzy model the paper described before and get the follow results. The fuzzy member is 0.86392, which means the average high speed is 365 km/h, the average medium high speed is 335 km/h, and headway time is 2.9319 minutes; there are also only one train late with the delay time between 0 minutes to 10 minutes and 2 trains late with the delay time between 10 minutes to 20 minutes; and the total delay time is 33.5 minutes. The adjustment strategy is to extend the dwell time for the train K101 and then make the train G105 overtake the train K101 at Changzhou North Station, which reflects the adjustment priority for high-grade train. Figure 3 shows detail rescheduling result by fuzzy model. The gray lines mean the initial timetable, and blue lines mean the rescheduling timetable.

4.1.2. Two Trains Delay

In the simulation, we assume that the train G305 is late for 30 minutes in the section from Nanjing to Zhenjiang, then the train G103 gets further delay for 20 minutes in the section from Zhenjiang to Changzhou under the condition of the existing delay, and other conditions are identical to Section 4.1.1.

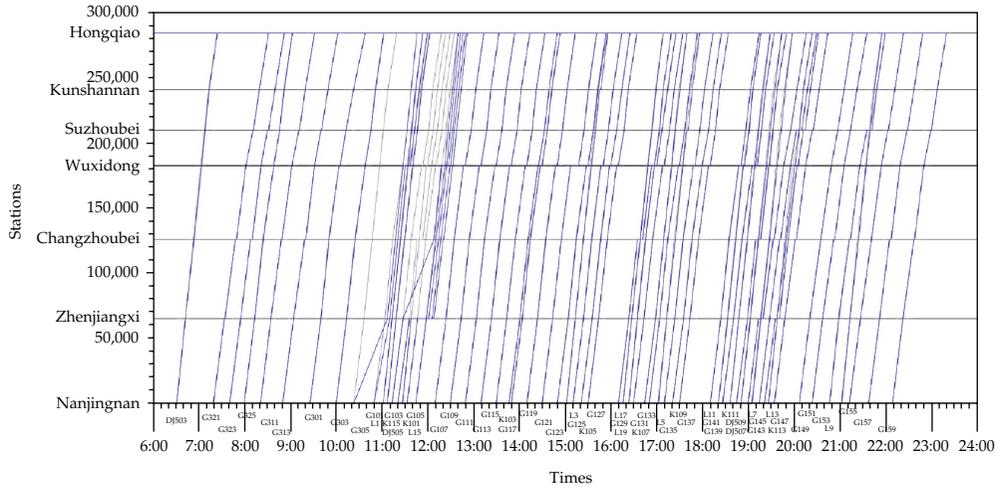


Figure 4: Timetable of two trains delay.

First we use the original model to solve the problem and get the follow results. When we take the strict constraints, the solution value is 860720; there are 2 trains late with the delay time between 0 minutes to 10 minutes, 3 trains late with the delay time between 10 minutes to 20 minutes, and 3 trains late with the delay time between 20 minutes to 30 minutes; the total delay time is 122.43 minutes. When we relax the constraints, the solution value is 848871; there are 2 trains late with the delay time between 0 minutes to 10 minutes, 2 trains late with the delay time between 10 minutes to 20 minutes, and also 3 trains late with the delay time between 20 minutes to 30 minutes; and the total delay time is 109.18 minutes. Then we solve the fuzzy model and get the follow results. The fuzzy member is 0.89425, which means the average high speed is 362 km/h, the average medium high speed is 332 km/h, and headway time is 2.9512 minutes; there are also 7 delay trains, 2 trains of which late with the delay time between 0 minutes to 10 minutes, 2 trains late with the delay time between 10 minutes to 20 minutes, and 3 trains late with the delay time between 20 minutes to 30 minutes; and the total delay time is 114.29 minutes. The adjustment strategy is to make the high level train G105 subsequently overtakes the train K115 L15 and K101 at Zhenjiang West Station, and train L15 overtakes the train G103 at the Wuxi East Station, which effectively avoid the high level train G105 being late and reduce the delay time of the train L15. Figure 4 shows detail rescheduling result by fuzzy model.

4.1.3. Speed Restriction in All Sections

When some natural hazards happened, like heavy storm and strong wind, railway will be greatly affected in a large area. In the simulation, we assume there is a speed restriction in all sections and the average limited speed ranges from 170 km/h to 150 km/h; other conditions are identical to Section 4.1.1.

First we use the original model to solve the problem and get the follow results. When we set the speed as 150 km/h and headway time as 3 minutes, the solution value is 7969213; there are 2 trains late for the time between 30 and 40 minutes, 18 trains late for the time between 40 and 50 minutes, 34 trains late for the time between 50 and 60 minutes, and

6 trains late over one hour; and the total delay time is 3052.95 minutes. When the speed is set to 170 km/h, and headway time is set to 2.5 minutes, the solution value is 7473849; there are 5 trains late for the time between 20 and 30 minutes, 38 trains late for the time between 30 and 40 minutes, 17 trains late for the time between 40 and 50 minutes, and no train late over one hour; and the total delay time is 2248.26 minutes. Although the latter result is very exciting, it is a significant risk to take the speed of 170 km/h, as the highest speed must bring the high operation cost and may arouse some new delays. Then we use the above results as the inputs of the fuzzy model the paper described before and get the follow results. The fuzzy member is 0.689117, which means the average speed is 156 km/h, and headway time is 2.8445 minutes; there are 19 trains late for the time between 30 and 40 minutes, 25 trains late for the time between 40 and 50 minutes, 16 trains late for the time between 50 and 60 minutes, and no train late over one hour; and the total delay time is 2655.16 minutes. We can see that the fuzzy optimization result improved greatly with little slack of constraints that means little risk and operation cost. Figure 5 is the rescheduling timetable by fuzzy model. In addition, the result is a parallel diagram since all trains must comply with the same speed.

4.1.4. Speed Restriction in One Section

Sometimes equipment failure, like train signal failure, happens in some but not all the sections. In the simulation, we assume there is a speed restriction in first section, and the limited speed ranges from 60 km/h to 50 km/h; other conditions are identical to Section 4.1.1.

First we use the original model to solve the problem and get the follow result. When we take the strict constraints, the solution value is 9294339; there are 19 trains late for the time between 40 and 50 minutes, 30 trains late for the time between 50 and 60 minutes, and 11 trains late over one hour; and the total delay time is 3269.32 minutes. When the speed is 60 km/h, and headway time is 2.5 minutes, the solution value is 8577661; there are 20 trains late for the time between 30 and 40 minutes, 37 trains late for the time between 40 and 50 minutes, 3 trains late for the time between 50 and 60 minutes, and no train late over one hour; and the total delay time is 2486.32 minutes. Then we get the follow result using the fuzzy model the paper described before. The fuzzy member is 0.70606, which means the average speed is 53 km/h, and headway time is 2.85303 minutes; there are 5 trains late for the time between 30 and 40 minutes, 39 trains late for the time between 40 and 50 minutes, 16 trains late for the time between 50 and 60, minutes and no train late over one hour; the total delay time is 2753.17 minutes. The adjustment strategy is to make the train G117 overtakes the train K105 at Wuxi East Station, and then let the train L7 overtake the train G143 at the Wuxi East Station. Figure 6 is the rescheduling timetable with speed restriction in the first section.

4.1.5. Speed Restriction in One Section with One Train Delay

In the simulation, we assume that there is a speed restriction in first section, and the limited speed ranges from 60 km/h to 50 km/h; the train G103 gets delay in the section from Zhenjiang to Changzhou. Other conditions are identical to Section 4.1.1.

First we use the original model to solve the problem and get the following results. When the speed in the first section is 50 km/h; the trains can run at the normal speed; the headway time is set to 3 minutes, the solution value is 9367637; there are 19 trains late with the delay time between 40 minutes to 50 minutes, 24 trains late with the delay time between 50 minutes to 60 minutes, and 17 trains late for more than one hour; and the total delay time

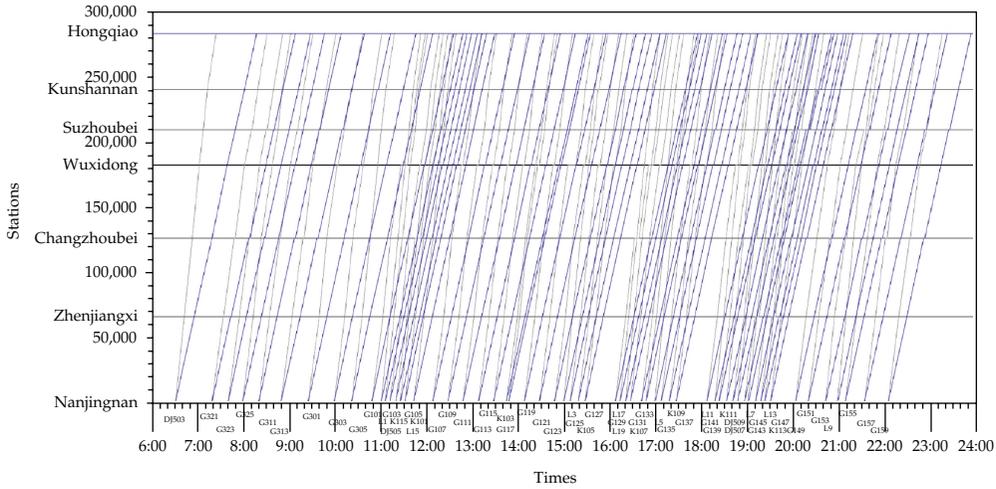


Figure 5: Timetable of speed restriction in all sections.

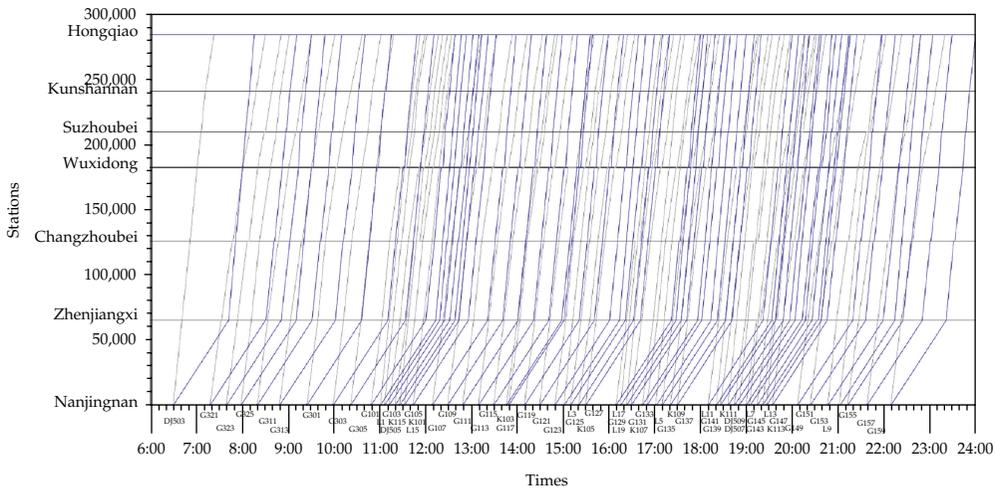


Figure 6: Timetable of speed restriction in first section.

is 3305.19 minutes. When the speed in the first section is 60 km/h; the headway time is set to 2.5 minutes; then the solution value is 8671072; 3 trains late with the delay time between 20 minutes to 30 minutes, 34 trains late with the delay time between 30 minutes to 40 minutes, 17 trains late with the delay time between 40 minutes to 50 minutes, 1 train late with the delay time between 50 minutes to 60 minutes, and 5 train late with the delay time more than one hour; and the total delay time is 2557.32 minutes. Then we use the above results as the inputs of the fuzzy model the paper described before and get the follow results. The fuzzy member is 0.740856, which means the average speed in the first section is 52 km/h, and headway time is 2.86 minutes; there are 17 trains late with the delay time between 30 minutes to 40 minutes, 23 trains late with the delay time between 40 minutes to 50 minutes, 15 trains late with the delay time between 50 minutes to 60 minutes, and 5 trains late with the delay time more than one hour; and the total delay time is 2823.34 minutes. The adjustment strategy is to make the

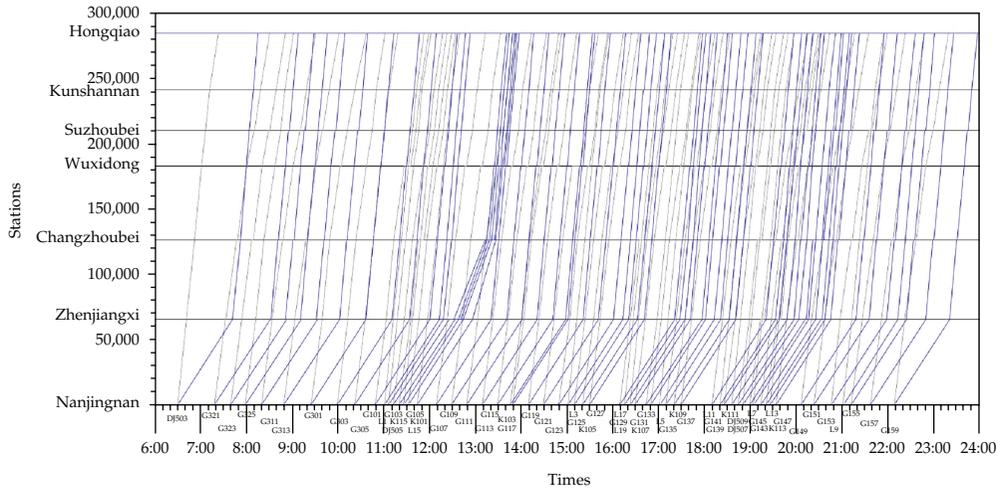


Figure 7: Timetable of speed restriction in one section and one train delay.

trains G105, K115, and L15 overtake the train K115 at Changzhou North Station, and then let the train G105 overtake the train K101 at the Wuxi East Station. Figure 7 is the rescheduling timetable by fuzzy model.

4.1.6. Speed Restriction in All Sections with Two Trains Delay

In the simulation, we assume that the train G305 is late for 40 minutes in the section between Nanjing to Zhenjiang and then get further delay for 15 minutes in the section between Zhenjiang to Changzhou; the train G103 gets a delay for 30 minutes in the section from Zhenjiang to Changzhou; also there is a speed restriction in all sections, and the average limited speed ranges from 170 km/h to 150 km/h for all the trains; other conditions are identical to Section 4.1.1.

First we use the original model to solve the problem and get the follow results. When we set the speed as 150 km/h and headway time as 3 minutes, the solution value is 8042152; there are 1 train late with the delay time between 30 minutes to 40 minutes, 17 trains late with the delay time between 40 minutes to 50 minutes, 27 trains late with the delay time between 50 minutes to 60 minutes, and 15 trains late with the delay time more than one hour; and the total delay time is 3197.86 minutes. When the speed is 170 km/h, and headway time is 2.5 minutes, the solution value is 7588658; there are 5 trains late with the delay time between 20 minutes to 30 minutes, 35 trains late with the delay time between 30 minutes to 40 minutes, 14 trains late with the delay time between 40 minutes to 50 minutes, 3 trains late with the delay time between 50 minutes to 60 minutes, and 3 trains late with the delay time more than one hour; and the total delay time is 2397.15 minutes. Then we use the above results as the inputs of the fuzzy model the paper described before and get the follow results. The fuzzy member is 0.792877, which means the average speed is 153 km/h, and headway time is 2.89 minutes; there are 17 trains late with the delay time between 30 minutes to 40 minutes, 23 trains late with the delay time between 40 minutes to 50 minutes, 15 trains late with the delay time between 50 minutes to 60 minutes, and 5 trains late with the delay time more than one hour; and the total delay time is 2704.43 minutes. The adjustment strategy is to make the train

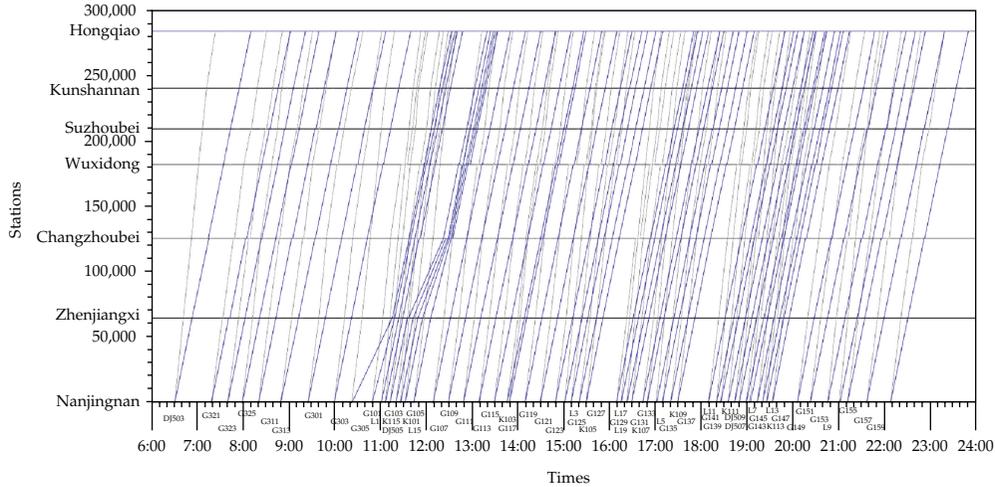


Figure 8: Timetable of speed restriction in all sections and two trains delay.

L1 overtakes the train G101 at the Wuxi East Station, the train L15 and the train K101 overtake K115, and then extend the dwell time of K115 at the Wuxi East Station for the stopover by the train G101, so the importance of the high level trains is also shown from the steps above. Figure 8 is the rescheduling timetable by fuzzy model.

These simulations are realized on the computer with Intel Core 2 Duo CPU E7500 and 2sG Memory. All the optimization results for the 6 cases are given in Table 1, where "R1" denotes original optimization subject to high speed, "R2" denotes original optimization subject to low speed, "R3" denotes fuzzy optimization, "D1" denotes the number of trains with a delay less than 10 mins, "D2" denotes the number of trains with a delay between 10 and 20 mins, "D3" denotes the number of trains with a delay between 20 and 30 mins, "D4" denotes the number of trains with a delay between 30 and 40 mins, "D5" denotes the number of trains with a delay between 40 and 50 mins, "D6" denotes the number of trains with a delay between 50 and 60 mins, "D7" denotes the number of trains with a delay not less than 60 mins, "D8" denotes the objective value, "D9" denotes the total delay time, "D10" denotes the number of stopover trains, and "D11" denotes the calculation time.

4.2. Different Weights

There are four kinds of fuzzy constraints as described in Sections 2.4 and 3.2. Since the minimum separation time on track possession in each station is set to 1 min in the previous experiments, we can ignore the uncertainty of station dwell time constraints. Furthermore, all the station and section headway times are set to the same value that from 3 minutes to 2.5 minutes; thereby the two constraints are considered equally important for the optimal objective and can use the same fuzzy member. Therefore two kinds of fuzzy constraints are analyzed in this section, section running time and headway time, and the weights of their fuzzy member are w_1 and w_2 .

Table 2 and Figure 9 are the comparison of the optimal objectives for the six cases. It can be seen from Figure 9 that the objective increases obviously as w_1 is set from 0.1 to 0.9 in the last four cases; however, the objective curves tend to flat in the former two cases,

Table 1: Rescheduling results of different cases.

Case	R	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
Case 1	R1	0	3	1	0	0	0	0	368784	47.5	—	5.38
	R2	1	2	0	0	0	0	0	350449	31.5	—	3.39
	R3	1	2	0	0	0	0	0	0.86392	33.5	1	178.17
Case 2	R1	2	3	3	0	0	0	0	860720	122.43	—	6.22
	R2	2	2	3	0	0	0	0	848871	109.18	—	3.36
	R3	2	2	3	0	0	0	0	0.89425	114.29	2	142.58
Case 3	R1	0	0	0	2	18	34	6	7969213	3052.95	—	3.38
	R2	0	0	5	38	17	0	0	7473849	2248.26	—	3.32
	R3	0	0	0	19	25	16	0	0.689117	2655.16	—	23.13
Case 4	R1	0	0	0	0	19	30	11	9294339	3269.32	—	3.36
	R2	0	0	0	20	37	3	0	8577661	2486.32	—	3.39
	R3	0	0	0	5	39	16	0	0.70606	2753.17	2	76.75
Case 5	R1	0	0	0	0	19	24	17	9367637	3305.19	—	8.05
	R2	0	0	3	34	17	1	5	8671072	2557.32	—	7.97
	R3	0	0	0	17	23	15	5	0.740856	2823.34	3	280.23
Case 6	R1	0	0	0	1	17	27	15	8086249	3197.86	—	7.74
	R2	0	0	5	35	14	3	3	7588658	2397.15	—	7.75
	R3	0	0	0	17	23	15	5	0.792877	2704.43	4	260.56

Table 2: Comparison of the optimal objectives.

Case	w_1									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Case 1	0.860180	0.860920	0.861390	0.862250	0.863920	0.864470	0.865731	0.865020	0.864170	
Case 2	0.889860	0.890440	0.891050	0.892280	0.894240	0.896030	0.895490	0.892661	0.890040	
Case 3	0.671015	0.674356	0.678921	0.684357	0.689117	0.693258	0.698091	0.702655	0.705186	
Case 4	0.697540	0.698920	0.700130	0.702480	0.706060	0.709138	0.712240	0.714521	0.715390	
Case 5	0.730050	0.732160	0.734875	0.737280	0.740856	0.744030	0.747980	0.749721	0.751060	
Case 6	0.779810	0.781530	0.784010	0.787903	0.792877	0.797460	0.801130	0.804655	0.806515	

especially the curves decline slightly when w_1 is greater than 0.6. This may explain that the same constraints contribute differently for the optimal objective in different emergencies. Section running time constraint has a higher priority than headway time constraint when some accidents or natural disasters lead to speed restriction for a long time, and adjusting the train headway time is more effective if few trains delay by some interference.

5. Conclusion and Future Work

This paper presents a fuzzy optimization model based on improved tolerance approach for train rescheduling in case of train delay and speed restriction, which deals with the train running time at sections, the headway time at sections, and station as the fuzzy parameters. The simulations on Beijing to Shanghai high speed line reveal that the fuzzy optimization result improved greatly with little slack of constraints. This means we can get a new timetable with less total delay time as well as the number of seriously impacted trains in safe and lower average speed, little dwell time, and enough headway time. In addition, the sensitivity

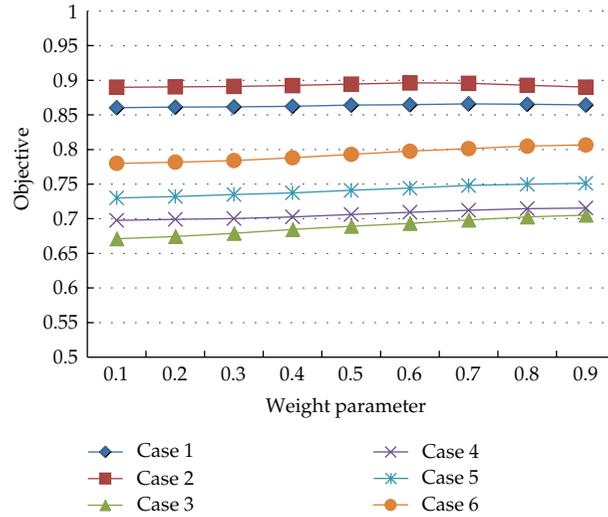


Figure 9: Optimal objectives with different weights.

analysis of the weighing factors shows that the same constraints contribute differently to the optimal objective in different emergencies, thus the dispatchers should take different trains adjustment strategies to eliminate interference as much as possible.

There remain many interesting areas to explore around the uncertainty in timetable rescheduling problem. Firstly, the membership functions of fuzzy parameters used in the paper may be more complex form in practice than the linear function in the paper, so we can take some genetic functions, like gauss membership function, to model the fuzzy programming. The more accurate the membership function is, the better result the fuzzy optimization model gets. Secondly, the tolerance can also be described by fuzzy set-based schemes. Finally, fuzzy operators in the model can be improved to adapt well to the information processing mechanism of despatchers in dealing with rescheduling problems. Our ultimate goal is to develop a real-time rescheduling system to significantly improve operation management and scheduling efficiency in the future.

Acknowledgments

This work has been supported by National Natural Science Foundation of China (61074151); National Key Technology Research and Development Program (2009BAG12A10); Research Found of State Key Laboratory of Rail Traffic Control and Safety (RCS2009ZT002, RCS2010ZZ002, and RCS2011ZZ004).

References

- [1] "Medium-Long Term Railway Network Plan (2008 Revision)," China Railway Ministry, 2008.
- [2] 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.
- [3] X. Zhou and M. Zhong, "Single-track train timetabling with guaranteed optimality: Branch-and-bound algorithms with enhanced lower bounds," *Transportation Research Part B*, vol. 41, no. 3, pp. 320–341, 2007.

- [4] F. Corman, A. D'Ariano, D. Pacciarelli, and M. Pranzo, "Optimal inter-area coordination of train rescheduling decisions," *Transportation Research Part E*, vol. 48, no. 1, pp. 71–88, 2012.
- [5] R. Cordone and F. Redaelli, "Optimizing the demand captured by a railway system with a regular timetable," *Transportation Research Part B*, vol. 45, no. 2, pp. 430–446, 2011.
- [6] 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.
- [7] R. M. P. Goverde, "Railway timetable stability analysis using max-plus system theory," *Transportation Research Part B*, vol. 41, no. 2, pp. 179–201, 2007.
- [8] P. Vansteenwegen and D. V. Oudheusden, "Developing railway timetables which guarantee a better service," *European Journal of Operational Research*, vol. 173, no. 1, pp. 337–350, 2006.
- [9] E. Castillo, I. Gallego, J. M. Ureña, and J. M. Coronado, "Timetabling optimization of a mixed double- and single-tracked railway network," *Applied Mathematical Modelling*, vol. 35, no. 2, pp. 859–878, 2011.
- [10] M. J. Dorfman and J. Medanic, "Scheduling trains on a railway network using a discrete event model of railway traffic," *Transportation Research Part B*, vol. 38, no. 1, pp. 81–98, 2004.
- [11] T. W. Chiang, H. Y. Hau, H. M. Chiang, S. Y. Ko, and C. H. Hsieh, "Knowledge-based system for railway scheduling," *Data and Knowledge Engineering*, vol. 27, no. 3, pp. 289–312, 1998.
- [12] A. Fay, "A fuzzy knowledge-based system for railway traffic control," *Engineering Applications of Artificial Intelligence*, vol. 13, no. 6, pp. 719–729, 2000.
- [13] F. Corman, A. D'Ariano, D. Pacciarelli, and M. Pranzo, "A tabu search algorithm for rerouting trains during rail operations," *Transportation Research Part B*, vol. 44, no. 1, pp. 175–192, 2010.
- [14] C. Z. Yin, L. Bu, X. Q. Cheng, and Y. Pu, "Tabu search algorithm on location for railway baggage and parcel base and distribution sites," *Control and Decision*, vol. 21, no. 11, pp. 1316–1320, 2006.
- [15] S. Q. Dong, J. Y. Wang, and H. F. Yan, "Tabu search for train operation adjustment on double-track line," *China Railway Science*, vol. 26, no. 4, pp. 114–119, 2005.
- [16] 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.
- [17] J. W. Chung, S. M. Oh, and I. C. Choi, "A hybrid genetic algorithm for train sequencing in the Korean railway," *Omega*, vol. 37, no. 3, pp. 555–565, 2009.
- [18] S. He, R. Song, and S. S. Chaudhry, "Fuzzy dispatching model and genetic algorithms for railyards operations," *European Journal of Operational Research*, vol. 124, no. 2, pp. 307–331, 2000.
- [19] P. Ren, N. Li, and L. Q. Gao, "Bi-criteria passenger trains scheduling optimal planning based on integrated particle swarm optimization," *Journal of System Simulation*, vol. 19, no. 7, pp. 1449–1452, 2007.
- [20] X. Meng, L. Jia, and Y. Qin, "Train timetable optimizing and rescheduling based on improved particle swarm algorithm," *Transportation Research Record*, no. 2197, pp. 71–79, 2010.
- [21] T. K. Ho, C. W. Tsang, K. H. Ip, and K. S. Kwan, "Train service timetabling in railway open markets by particle swarm optimisation," *Expert Systems with Applications*, vol. 39, no. 1, pp. 861–868, 2012.
- [22] F. Zhao and X. Zeng, "Optimization of transit route network, vehicle headways and timetables for large-scale transit networks," *European Journal of Operational Research*, vol. 186, no. 2, pp. 841–855, 2008.
- [23] A. Jamili, M. A. Shafia, S. J. Sadjadi, and R. Tavakkoli-Moghaddam, "Solving a periodic single-track train timetabling problem by an efficient hybrid algorithm," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 4, pp. 793–800, 2012.
- [24] S. Mu and M. Dessouky, "Scheduling freight trains traveling on complex networks," *Transportation Research Part B*, vol. 45, no. 7, pp. 1103–1123, 2011.
- [25] X. Zhou and M. Zhong, "Bicriteria train scheduling for high-speed passenger railroad planning applications," *European Journal of Operational Research*, vol. 167, no. 3, pp. 752–771, 2005.
- [26] A. Caprara, M. Monaci, P. Toth, and P. L. Guida, "A Lagrangian heuristic algorithm for a real-world train timetabling problem," *Discrete Applied Mathematics*, vol. 154, no. 5, pp. 738–753, 2006.
- [27] W. X. Zha and G. Xiong, "Study on the dynamic performance of the train running time in the section," *Systems Engineering*, vol. 19, no. 1, pp. 47–51, 2001.
- [28] X. C. Zhang and A. Z. Hu, "Analysis of β -function distribution for deviation of train running time in section," *Journal of the China Railway Society*, vol. 18, no. 3, pp. 1–6, 1996.

- [29] L. M. Jia and X. D. Zhang, "Distributed intelligent railway traffic control based on fuzzy decisionmaking," *Fuzzy Sets and Systems*, vol. 62, no. 3, pp. 255–265, 1994.
- [30] L. Yang, K. Li, and Z. Gao, "Train timetable problem on a single-line railway with fuzzy passenger demand," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 3, pp. 617–629, 2009.
- [31] L. Yang, Z. Gao, and K. Li, "Railway freight transportation planning with mixed uncertainty of randomness and fuzziness," *Applied Soft Computing Journal*, vol. 11, no. 1, pp. 778–792, 2011.
- [32] M. A. Shafia, S. J. Sadjadi, A. Jamili, R. Tavakkoli-Moghaddamb, and M. Pourseyed-Aghaeec, "The periodicity and robustness in a single-track train scheduling problem," *Applied Soft Computing*, vol. 12, no. 1, pp. 440–452, 2012.
- [33] 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.
- [34] W. H. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, Communications Press, Beijing, China, 2012.
- [35] W. H. Wang, W. Zhang, H. W. Guo, H. Bubb, and K. Ikeuchi, "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research Part C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [36] H. W. Guo, W. H. Wang, W. W. Guo, X. B. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [37] L. Wang, L. M. Jia, Y. Qin, J. Xu, and W. 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.
- [38] X. M. Shi, "Research on train headway time of high speed railway in China," *Chinese Railways*, no. 5, pp. 32–35, 2005.
- [39] C. V. Negiata and M. Sularia, "On fuzzy mathematical programming and tolerances in planning," *Economic Computation and Economic Cybernetics Studies and Research*, no. 1, pp. 3–15, 1976.
- [40] H. Rommelfanger, "Fuzzy linear programming and applications," *European Journal of Operational Research*, vol. 92, no. 3, pp. 512–527, 1996.
- [41] B. Werners, "An interactive fuzzy programming system," *Fuzzy Sets and Systems*, vol. 23, no. 1, pp. 131–147, 1987.
- [42] S. C. Fang and D. W. Wang, *Fuzzy Math and Fuzzy Optimization*, Science Press, Beijing, China, 1997.

Research Article

Lane-Level Vehicle Trajectory Reckoning for Cooperative Vehicle-Infrastructure System

Yinsong Wang, Xiaoguang Yang, Luoyi Huang, and Jiawen Wang

*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 Yinsong Wang, piny-song@163.com

Received 11 September 2012; Revised 26 November 2012; Accepted 27 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Yinsong 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 presents a lane-level positioning method by trajectory reckoning without Global Positioning System (GPS) equipment in the environment of Cooperative Vehicle-Infrastructure System (CVIS). Firstly, the accuracy requirements of vehicle position in CVIS applications and the applicability of GPS positioning methods were analyzed. Then, a trajectory reckoning method based on speed and steering data from vehicle's Control Area Network (CAN) and roadside calibration facilities was proposed, which consists of three critical models, including real-time estimation of steering angle and vehicle direction, vehicle movement reckoning, and wireless calibration. Finally, the proposed method was validated through simulation and field tests under a variety of traffic conditions. Results show that the accuracy of the reckoned vehicle position can reach the lane level and match the requirements of common CVIS applications.

1. Introduction

The recently developed Cooperative Vehicle-Infrastructure System (CVIS) enables vehicle with the capability to communicate with other vehicles and infrastructure and to ensure safety and travel efficiency by driving assistance [1]. The capability of providing accurate position estimation is essential for successful and widespread deployment of CVIS applications such as optimal speed advisory and lane change assistant.

The Global Positioning System (GPS) is one of the most convenient and accurate methods for determining vehicle position in a global coordinate system [2, 3]. Low-cost GPS receivers can provide positioning service with an accuracy of approximately 10–20 meters, and the Differential GPS (DGPS) receivers have improved the positioning accuracy to approximately 2–3 meters, but require high construction cost for both onboard receiver and base station. To realize lane-level positioning, vehicle movement dynamics was integrated

into the position estimate by an Extended Kalman filter utilizing only low-cost GPS equipment, and the GPS errors were estimated from an event-driven vehicle turn and lane-alignment optimization model based on an accurate GPS description of the roadway [4]. In addition, a vehicle-lane-determining system was described in [5], which consists of an onboard DGPS receiver connecting with a wireless communications channel, a unique lane-level digital roadway database, a developed lane-match algorithm, and a real-time vehicle location display. Deduced Reckoning (commonly referred to as “Dead” Reckoning or DR) sensors consisting of an odometer and a gyroscope are routinely used to bridge any gaps in GPS positioning [6]. This information is then used with spatial road network data to determine the spatial reference of vehicle location via a process known as map matching [7]. Toledo-Moreo presented a solution that integrated a Global Navigation Satellite System (GNSS) receiver, an odometer, and a gyroscope, along with the road information stored in enhanced digital maps to analyze the reckoning trajectory [8].

On the other hand, most of the CVIS applications do not need the vehicle position within a global coordinate system. Positioning methods in relative coordinate systems such as Radio Frequency Identification (RFID) were developed recently. Accurate indoor position estimation utilizing passive RFID has been realized [9, 10]. In road traffic, Lee proposed a RFID-assisted localization system, which employed the DGPS concept to improve GPS accuracy [11]. Scheme of active RFID positioning of vehicles was developed in [12], the RFID tags were installed along lanes on road to form RFID matrix and the vehicle equipped with RFID readers locate itself by retrieving position information from tags around it.

This paper explores lane-level positioning by vehicle trajectory reckoning utilizing speed and steering data from vehicle’s Control Area Network (CAN). In addition, since the method was deployed in CVIS environment, the roadside calibration facility was designed by utilizing Detected Short Range Communication (DSRC) technology to improve the trajectory reckoning.

2. Accuracy Requirements for Positioning

2.1. The CVIS Environment

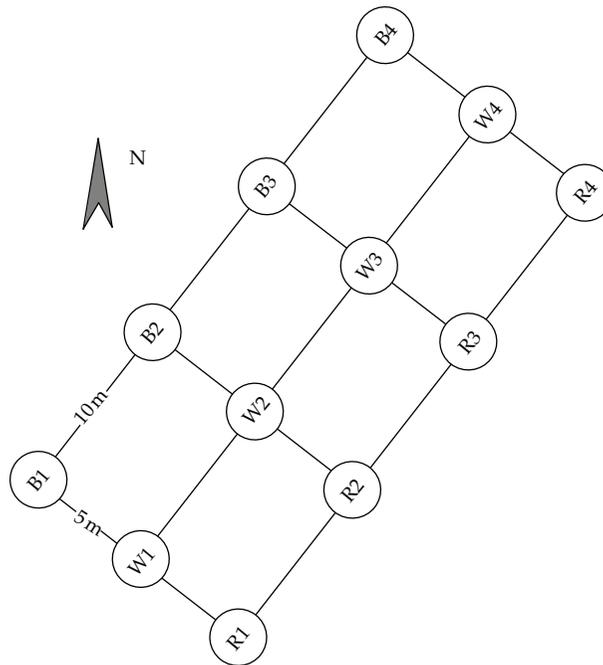
In a previous publication [13], a proof of concept for a prototype of Cooperative Vehicle-Infrastructure System (TJ-CVIS) developed at the University of Tongji was presented, which provides a vehicle with the capability to communicate with other vehicles and infrastructure. Conceptually, by using sensor technologies and standardized communication protocol, the vehicle probe data such as position, velocity, and preplanned routes can be automatically collected with high frequency and accuracy in CVIS. Therefore, a series of advanced traffic control applications and advanced driving assist applications for safety and efficiency can be deployed relying on the high sensitivity and information interactive abilities of CVIS. Table 1 lists the most common CVIS applications and their accuracy requirements for positioning [14].

2.2. GPS Position Errors

Low-cost GPS equipment has been widely used in the navigation applications. However, as the analysis shown in Section 2.1, the accuracy requirements of navigation and CVIS

Table 1: Typical CVIS applications and positioning accuracy requirements.

Application	Location/direction	Accuracy requirements
Route navigation	Vehicle heading	Road section level
Optimal speed advisory	Vehicle heading	Lane level
Automatic cruise	Vehicle heading	Lane level
Lane change assist	Lane heading	Lane level
Overtaking assist	Lane heading	Lane level
Collision avoidance	At intersection	Lane level
Active traffic control	At intersection	Lane level

**Figure 1:** Actual shape of test grid.

applications are markedly different. In order to test the feasibility of low-cost GPS equipment in CVIS, a common GPS receiver was placed at a set of stationary positions which form a grid shown in Figure 1. The positioning result is shown in Figure 2.

In Figure 2, the test result is totally out of shape compared to the actual shape of test sites. The distance between any two points in a line or row was calculated by the GPS data. The result shown in Table 2 indicates that in the situation of stationary positioning, the GPS errors arrive at 2.79 m on average, which is inapplicable for the CVIS applications.

In order to test the dynamic positioning accuracy of GPS, we conducted a test on an urban road, the starting and completing time stamps of lane changing were manually recorded. Figure 3 shows an example of dynamic positioning test in the situation of lane change. Results show that the performance of dynamic positioning is better than static

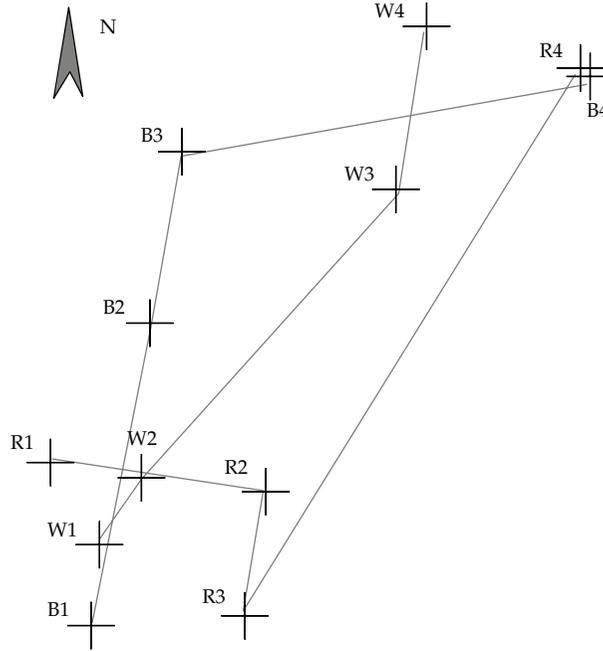


Figure 2: Stationary test result.

Table 2: Stationary positioning accuracy of GPS.

From	To	Calculated distance (m)	Actual distance (m)	Error (m)	Proportional error
B1	B2	27.12	10.00	-17.12	171.16%
B2	B3	7.47	10.00	2.53	25.25%
B3	B4	6.57	10.00	3.43	34.35%
B4	B1	23.14	30.00	6.86	22.86%
W1	W3	21.85	20.00	-1.85	9.25%
W1	W4	30.22	30.00	-0.22	0.74%
W3	W2	16.26	10.00	-6.26	62.57%
W4	W1	30.22	30.00	-0.22	0.74%
R1	R2	10.88	10.00	-0.88	8.83%
R2	R3	13.48	10.00	-3.48	34.84%
R3	R4	26.99	10.00	-16.99	169.90%
R4	R1	31.28	30.00	-1.28	4.27%
R4	R3	10.90	10.00	-0.90	9.04%

positioning [13]. Although map-matching can improve the positioning accuracy if good spatial road network data are available [15], a poor-quality road map could lead to a large error in map-matched solutions. Hence the lane changing event still cannot be easily recognized in the normal traffic environment.



Figure 3: Dynamic positioning test of GPS.

3. Vehicle Trajectory Reckoning

3.1. Input Parameters

The idea of lane-level positioning in this paper is reckoning the vehicle trajectory based on vehicle's driving-related data, then determining the position of vehicles. Figure 4(a) shows the actual vehicle trajectory of a left lane change event, Figures 4(b) and 4(c) show the variations of steering wheel angle and the front wheels' steering angle corresponding to the lane change event described in Figure 4(a). According to the data from CAN, the speed, steering wheel angle, and the wheels' steering angle can be recorded with a high frequency of 100 Hz, which enables the possibility to reckon the trajectory of a moving vehicle. Therefore, the vehicle position can be determined by the precise trajectory.

Since vehicle's body shape bears on the reckoning input, it is necessary to define the reference point of the vehicle. In this study, we developed the trajectory reckoning algorithm based on the four-wheel vehicle with a wheel track of B and a wheel base of L shown in Figure 5. For the four-wheel vehicle, each front wheel should perform only rolling movement around the rotation center O when steering. As an example of right steering shown in Figure 5, steering angle of the inner wheel (β) should be greater than the outer wheel (α). The relationship between β and α is described as the following:

$$\cot \alpha = \cot \beta + \frac{B}{L}. \quad (3.1)$$

To simplify the algorithm, we use the point of rotation axis of left-front wheel as the reference point shown as P in Figure 5. Finally, the input parameters can be specified as the vehicle body shape parameters, left-front wheel's speed (v), and its steering angle (α). For the vehicles which can only output the steering wheel angle (γ), α can be calculated by $\alpha = f(\gamma)$, which is specified by the vehicle type. Figure 6 shows an example of $f(\gamma)$ tested by a common four-wheel vehicle.

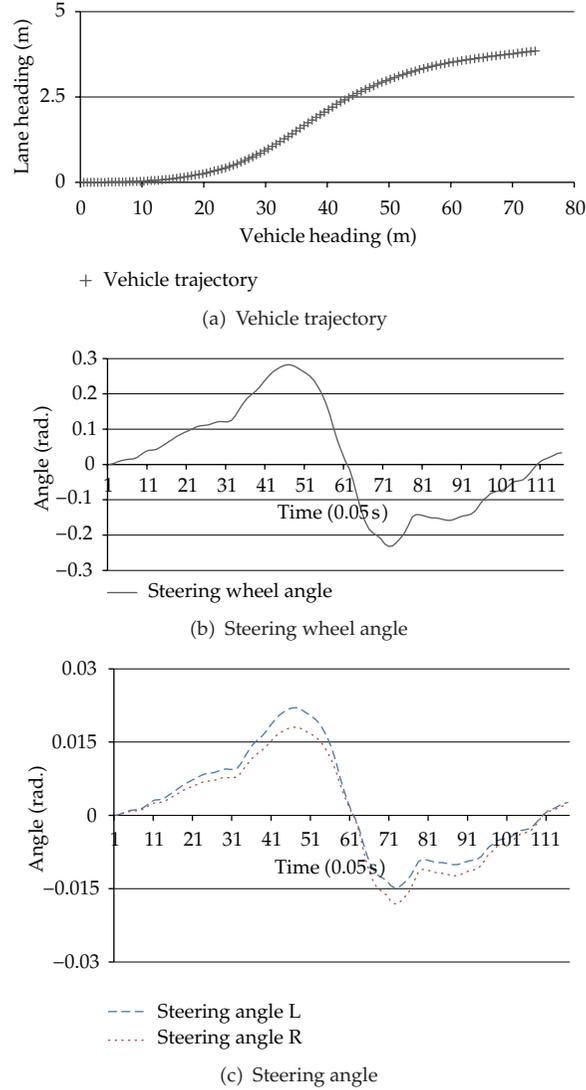


Figure 4: Variations of parameters in lane change event.

3.2. Trajectory Reckoning

Since GPS is not utilized in our test, which means global coordinate system is not available for the trajectory reckoning method, we choose the dynamic-update coordinate system shown in Figure 7. The calibration method will be introduced in Section 3.3.

At any time, the rotation radius (R , in meters—m) of the reference point can be calculated as the following:

$$R = \begin{cases} \frac{L}{\sin(|\alpha|)}, & |\alpha| > 0 \\ \infty, & |\alpha| = 0. \end{cases} \quad (3.2)$$

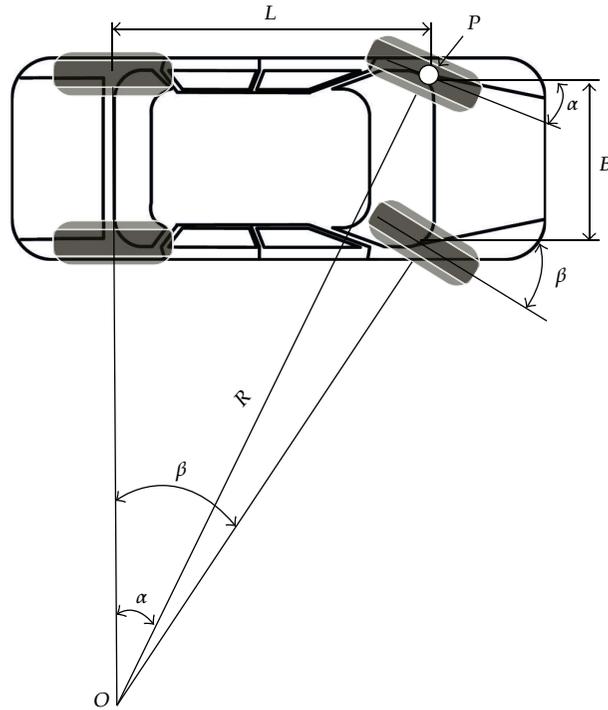


Figure 5: Optimal relationship of steering angle.

Angular speed (μ , in radian/second—rad/s) of reference point as the following:

$$\mu = \begin{cases} \frac{v}{R'}, & \alpha > 0 \\ 0, & \alpha = 0 \\ -\frac{v}{R'}, & \alpha < 0. \end{cases} \quad (3.3)$$

In (3.3), v is the linear speed of the reference point (in meters/second—m/s). When the trajectory is being reckoning, v and α are sampled with high frequency. Thus, the rotation angle (ω , in radian—rad) of the vehicle body can be calculated as the following:

$$\omega = \sum \mu_i \times t. \quad (3.4)$$

In (3.4), μ_i denotes the angular speed calculated by (3.3) at time sequence i , and t is the sampling interval.

The movements refer to the calibration point as:

$$\begin{aligned} \Delta x_i &= v_i \times \cos \omega \times t, \\ \Delta y_i &= v_i \times \sin \omega \times t. \end{aligned} \quad (3.5)$$

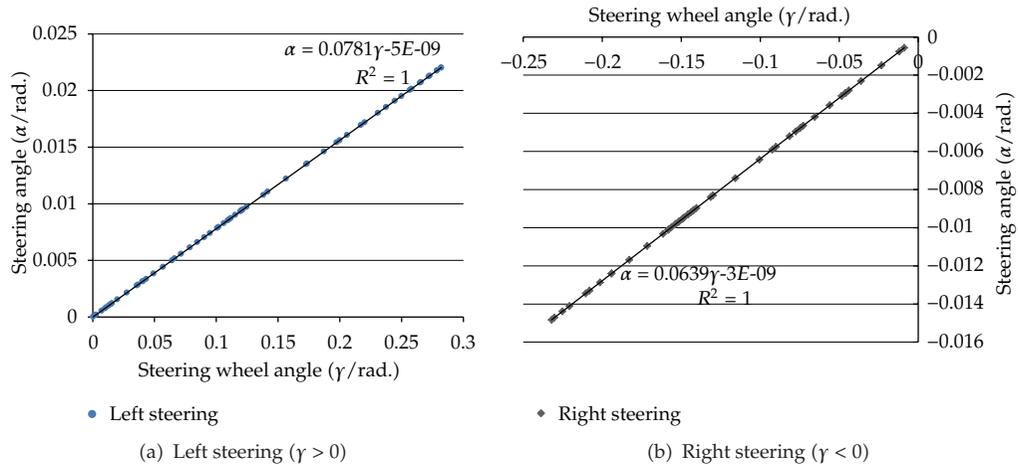
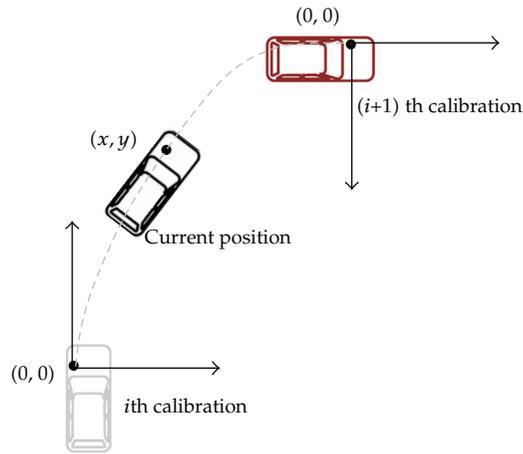
Figure 6: Example of $f(\gamma)$.

Figure 7: Dynamic update coordinate system.

In (3.5), v_i denotes the speed output at time sequence i , and the vehicle position at time sequence i can be reckoned as follows:

$$\begin{aligned} x_i &= \sum_{j=0}^i \Delta x_j, \\ y_i &= \sum_{j=0}^i \Delta y_j. \end{aligned} \tag{3.6}$$

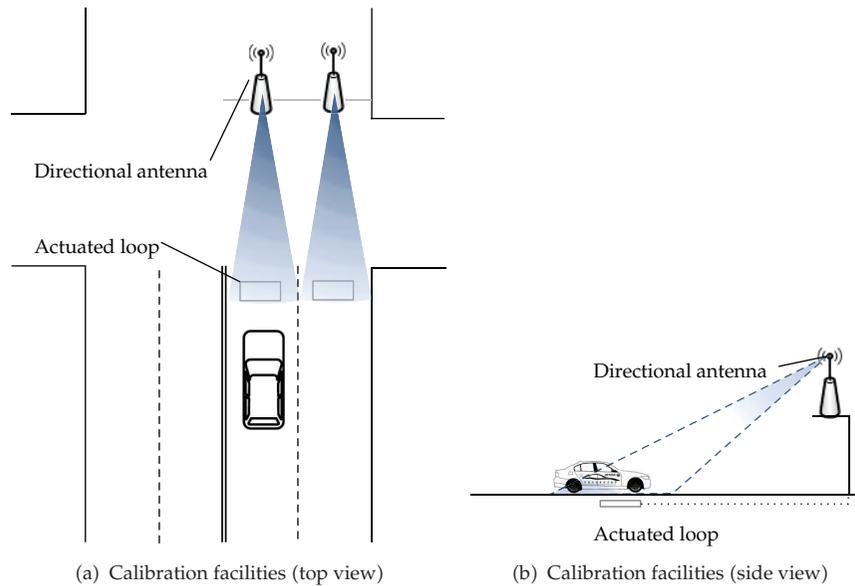


Figure 8: Calibration facilities.

3.3. Calibration Method

For the data output from CAN, the random errors cannot be avoided, and the cumulative errors in reckoning results increased rapidly when the distances get longer. It is necessary to calibrate the trajectory at a certain distance. Once the trajectory is calibrated, the dynamic-updated coordinate system referred above is updated.

The CVIS environment provides a vehicle with the capability to communicate with roadside infrastructures, which can be adopted to calibrate the vehicle trajectory. Figure 8 indicates the design of calibration facilities. The actuated loop, which has been already installed in most signalized intersections, is utilized in this method to detect the passing of vehicle. The directional antenna is set to send the calibration information to the vehicle by DSRC when vehicle is passing through the loop [16]. The road geometry data including road curvature, intersection geometry and lane drop geometry are stored in the onboard unit. In the prototype of TJ-CVIS [13], the content of calibration information was specified as the following:

- actuate time;
- road section ID;
- loop ID;
- lane ID of the loop;
- distance to the downstream calibration points.

With the calibration information, vehicle's position can be determined by comparing the reckoned trajectory with road geometry data. We assume that the vehicle is not



Figure 9: Field test Scenario 1.

steering when passing a loop at intersections. The calibration procedure is described as the following.

Step 1. Calibration message received by the vehicle.

Step 2. Update coordinate system referring to loop ID, lane ID and road section ID,

Step 3. $x_i = v_i \times t_{\text{latency}}$, $y_i = 0$, $\omega = 0$. (t_{latency} denotes the constant value of wireless communication latency, $t_{\text{latency}} = 0.01$ s in the prototype of TJ-CVIS [13]).

Step 4. Keep trajectory reckoning until next calibration.

4. Field and Driving Simulator Trials

4.1. Field Test

In order to validate the trajectory reckoning method, we conducted a field test at the campus of Tongji University. A computer program was developed on Windows platform to collect output data from CAN and reckon the trajectory. A common four-wheel vehicle was utilized as the experiment vehicle. Since the data protocol of CAN has not been disclosed by the manufacturer, we made use of the On-Board Diagnostics (OBD) protocol to read the vehicle speed. Hence only the algorithm for vehicle heading movement was validated in field test.

The first scenario of field test was deployed at a flattened square at the campus shown in Figure 9. The experiment vehicle was driven to pass through a pre-measured road segment with 37.5 m long. Another scenario of field test was deployed at a 276.6 m-long road shown in Figure 10. When experiment vehicle entering the test segment, the on-board unit started to reckon the trajectory. And when experiment vehicle leaving test segment, the distances of



Figure 10: Field test Scenario 2.

Table 3: Result of field test Scenario 1 (5 trials).

ID	Actual distance (m)	Reckoned distance (m)	Error (m)	PE
1	37.50	37.29	-0.21	0.56%
2	37.50	37.56	0.06	0.16%
3	37.50	37.83	0.33	0.88%
4	37.50	37.33	-0.17	0.45%
5	37.50	37.33	-0.17	0.45%

Table 4: Result of field test Scenario 2 (10 trials).

Actual distance (m)	Average reckoned distance (m)	ME (m)	MPE
276.60	277.54	-0.94	0.34%

trajectory between the vehicle's current position and the segment start line were recorded. The sampling frequency is 5 Hz in both scenarios. Test results are shown in Tables 3 and 4.

Results from field test indicate that the trajectory reckoning method can reach the submeter accuracy both in the two scenarios. It can also certify that the data from CAN are applicable for the reckoning method. Since the speed accuracy output by OBD protocol is limited to ± 1 km/h (0.28 m/s), the reckoning accuracy will be improved if we use the standard CAN protocol, which has a nominal output speed accuracy of ± 0.01 km/h (0.003 m/s).

4.2. Driving Simulator Trials

Another validation was taken on driving simulator due to the lack of CAN communication protocol of real vehicles. The driving simulator involved in validation was constructed by Tongji University, including a movement system with eight degrees of freedom shown in Figure 11. A real four-wheel vehicle with real steering system was installed in the movement system to simulate the situation of steering, braking, and acceleration. By using the driving simulator, the steering angle and speed can be sampled with high accuracy. Moreover, the actual trajectory can also be recorded for every sampling interval shown in Figure 12, which



Figure 11: Driving simulator in Tongji University.

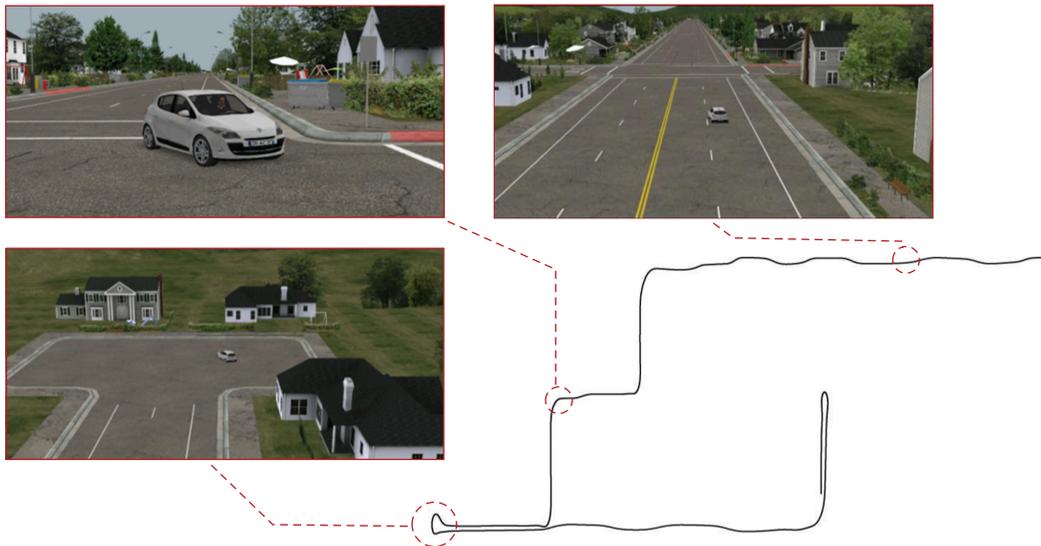
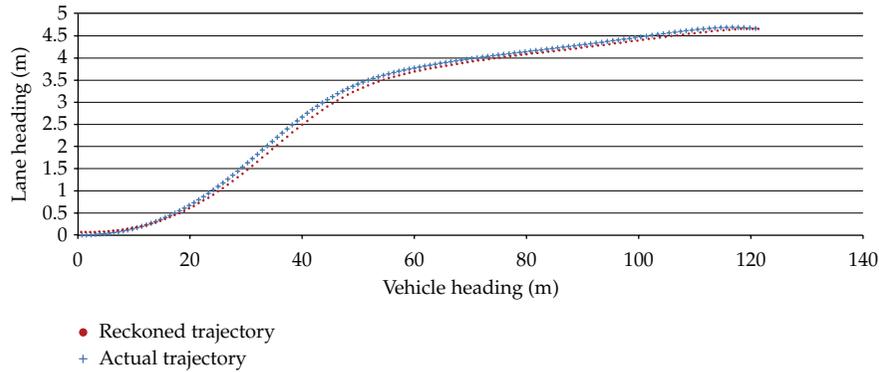


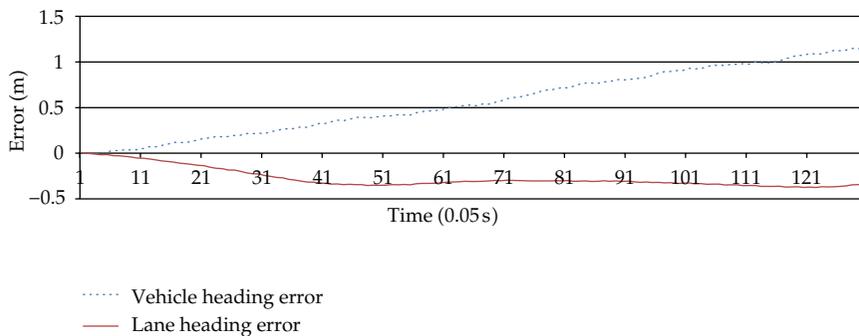
Figure 12: Actual trajectory and driving scenarios in driving simulator.

is hardly to record in field test and can make it easier to analyze the trajectory reckoning errors.

In the driving simulator trials, the driver was asked to change lane and turn at different speeds. The speed and steering angle of front-left wheel were recorded every 0.05 s. According to the manufactory, the wheel base (L) is 2.686 m. In addition, virtual calibration points were set at outbound lanes of each intersection. Examples of lane change trajectory and turning trajectory reckoning are shown in Figures 13 and 14.



(a) Trajectory reckoning result



(b) Reckoning error

Figure 13: Example of trajectory reckoning result of lane change.

The two examples indicate that the reckoning results are very close to the actual trajectories. If the geometry of the road and lane drop are known, the position of vehicle (i.e., distance to the next stop line and in which lane the vehicle is driving) can be determined by the reckoned trajectory.

Table 5 shows the statistical result of driving simulator trials, indicating that the proposed trajectory reckoning method can reach the accuracy of submeter both in vehicle heading and lane heading directions (U turn is not included in the statistic).

5. Conclusions and Future Work

This paper introduced a lane-level positioning method by trajectory reckoning without GPS equipment in the environment of CVIS. Vehicle driving data from CAN was adopted to reckon the vehicle movements, and then the position was determined by the reckoned trajectory object to the road geometry. In addition, a calibration method was designed to eliminate the cumulate errors during long-time reckoning. Results from field tests and driving simulator indicate that the proposed method can reach the positioning accuracy of submeter, which meets the requirements of most CVIS applications. Hence, the proposed method can be further deployed in the real traffic.

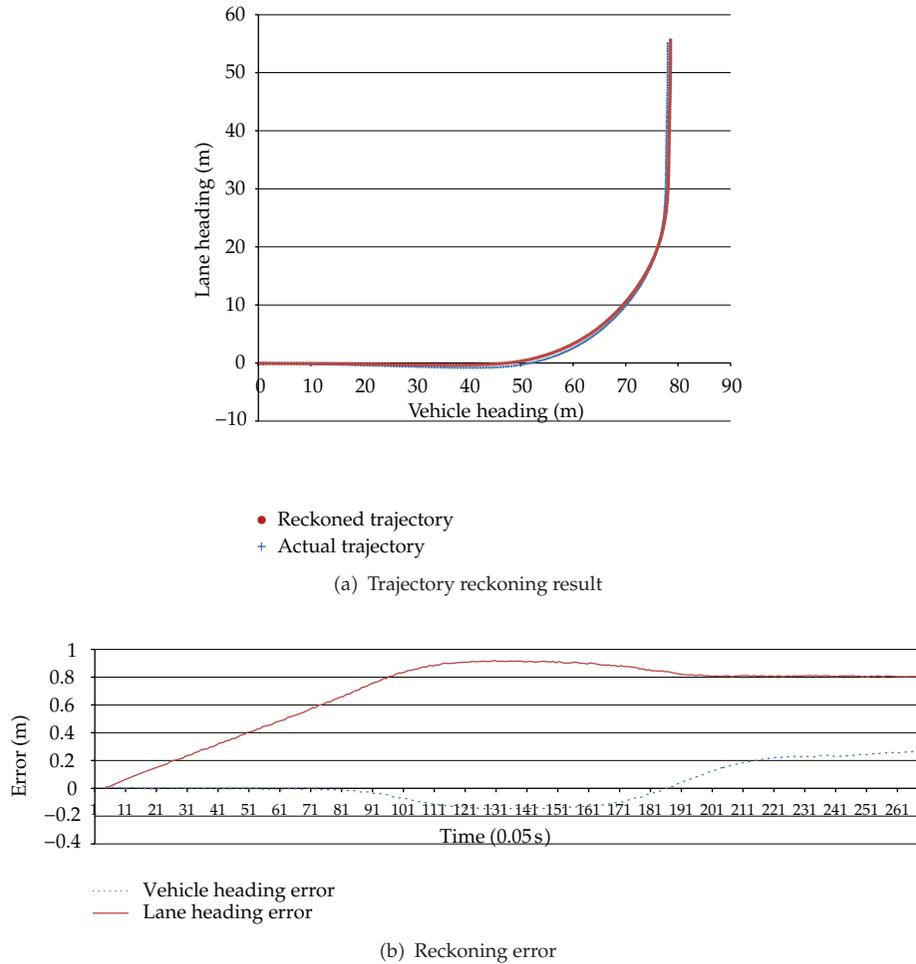


Figure 14: Example of trajectory reckoning result of turning.

Table 5: Trajectory reckoning results.

Event	Number of trials	MAE of x (vehicle heading, m)	MAE of y (lane heading, m)
Lane change	13	0.56	0.27
Turning	5	0.03	0.58

Although the method outperforms GPS, based on the field test and driving simulator trials, it has certain limitations. For instance, the system errors increase as the increase of length of trajectory takes place, and the performance of calibration facility drops when too many vehicles approach, which deserve further discussion.

Acknowledgments

This paper was made possible through funding from the Ministry of Science and Technology of China under Grant 2011AA110404 and the National Natural Science Foundation of China

under Grant 60974093. And the authors wish to acknowledge the vast and helpful assistance from Cooperative Vehicle-Infrastructure System Lab and the Driver Behavior and Traffic Safety Simulation Platform (Tongji University, China) for data collection.

References

- [1] W. H. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, China Communications Press, Beijing, China, 2012.
- [2] J. Du, J. Masters, and M. Barth, "Lane-level positioning for in-vehicle navigation and automated vehicle location (AVL) systems," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (ITSC '04)*, pp. 35–40, Washington, DC, USA, October 2004.
- [3] A. Boukerche, H. A. B. F. Oliveira, E. F. Nakamura, and A. A. F. Loureiro, "Vehicular Ad Hoc networks: a new challenge for localization-based systems," *Computer Communications*, vol. 31, no. 12, pp. 2838–2849, 2008.
- [4] Q. He and K. Larry Head, *Lane Level Vehicle Positioning With Low Cost GPS*, Transportation Research Board, Washington, DC, USA, 2010.
- [5] J. Du and M. J. Barth, "Next-generation automated vehicle location systems: positioning at the lane level," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 48–57, 2008.
- [6] D. Kubrak, C. Macabiau, M. Monnerat, and M. L. Bouchert, "Vehicular navigation using a tight integration of aided-GPS and low-cost MEMS sensors," in *Proceedings of the Institute of Navigation, National Technical Meeting (NTM '06)*, pp. 149–158, January 2006.
- [7] M. A. Quddus, W. Y. Ochieng, and R. B. Noland, "Current map-matching algorithms for transport applications: state-of-the art and future research directions," *Transportation Research C*, vol. 15, no. 5, pp. 312–328, 2007.
- [8] R. Toledo-Moreo, D. Bétaille, and F. Peyret, "Lane-level integrity provision for navigation and map matching with GNSS, dead reckoning, and enhanced maps," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 100–112, 2010.
- [9] N. Uchitomi, A. Inada, M. Fujimoto, T. Wada, K. Mutsuura, and H. Okada, "Accurate indoor position estimation by Swift-Communication Range Recognition (S-CRR) method in passive RFID systems," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '10)*, Zürich, Switzerland, September 2010.
- [10] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 37, no. 6, pp. 1067–1080, 2007.
- [11] E. K. Lee, S. Y. Oh, and M. Gerla, "RFID assisted vehicle positioning in VANETs," *Pervasive and Mobile Computing*, vol. 8, no. 2, pp. 167–179.
- [12] E. Zhang, W. Jiang, Y. Kuang et al., "Active RFID positioning of vehicles in road Traffic," in *Proceedings of the 11th International Symposium on Communications & Information Technologies (ISCIT '11)*, Hangzhou, China.
- [13] X. G. Yang, L. Y. Huang, Y. S. Wang, R. H. Du, J. W. Wang, and F. Yang, *A prototype of a cooperative vehicle infrastructure system: proof of concept—case study in Tongji University*, Transportation Research Board, Washington, DC, USA, 2012.
- [14] S. E. Shladover and S. K. Tan, "Analysis of vehicle positioning accuracy requirements for communication-based cooperative collision warning," *Journal of Intelligent Transportation Systems*, vol. 10, no. 3, pp. 131–140, 2006.
- [15] W. Y. Ochieng, M. A. Quddus, and R. B. Noland, "Map-matching in complex urban road networks," *Brazilian Journal of Cartography*, vol. 55, no. 2, pp. 1–18, 2004.
- [16] L. Sun, Y. Wu, J. Xu et al., "An RSU-assisted localization method in non-GPS highway traffic with dead reckoning and V2R communications," in *Proceedings of the 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet '12)*, Yichang, China, 2012.

Research Article

Analysis of ITMS System Impact Mechanism in Beijing Based on FD and Traffic Entropy

Ailing Huang, Wei Guan, Yimei Chang, and Zhen Yang

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Ailing Huang, alhuang@bjtu.edu.cn

Received 20 September 2012; Revised 22 November 2012; Accepted 27 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Ailing Huang 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.

Although more attention has been attracted to benefit evaluation of Intelligent Transportation Systems (ITS) deployment, how ITS impact the traffic system and make great effects is little considered. As a subsystem of ITS, in this paper, Intelligent Transportation Management System (ITMS) is studied with its impact mechanism on the road traffic system. Firstly, the correlative factors between ITMS and the road traffic system are presented and 3 positive feedback chains are defined. Secondly, we introduce the theory of Fundamental Diagram (FD) and traffic system entropy to demonstrate the correlative relationship between ITMS and feedback chains. The analyzed results show that ITMS, as a negative feedback factor, has damping functions on the coupling relationship of all 3 positive feedback chains. It indicates that with its deployment in Beijing, ITMS has impacted the improvement of efficiency and safety for the road traffic system. Finally, related benefits brought by ITMS are presented corresponding to the correlative factors, and effect standards are identified for evaluating ITMS comprehensive benefits.

1. Introduction

1.1. Background

Intelligent Transportation Management System (ITMS) is an advanced system for traffic control, traffic management, and decision-making deployed by road traffic administration. It utilizes technologies and methods of Intelligent Transportation Systems (ITS) to manage the urban road traffic system efficiently, systematically and scientifically, to ensure the traffic system is safer and traffic flow is smoother, promote the environmental protection, and realize energy resource saving. ITMS is one of the research fields of ITS and is one type of ITS applications in road traffic management. The ITMS of Beijing was started to deploy in 2004 and was established and put into practice entirely in 2008. It consists of 1 center and

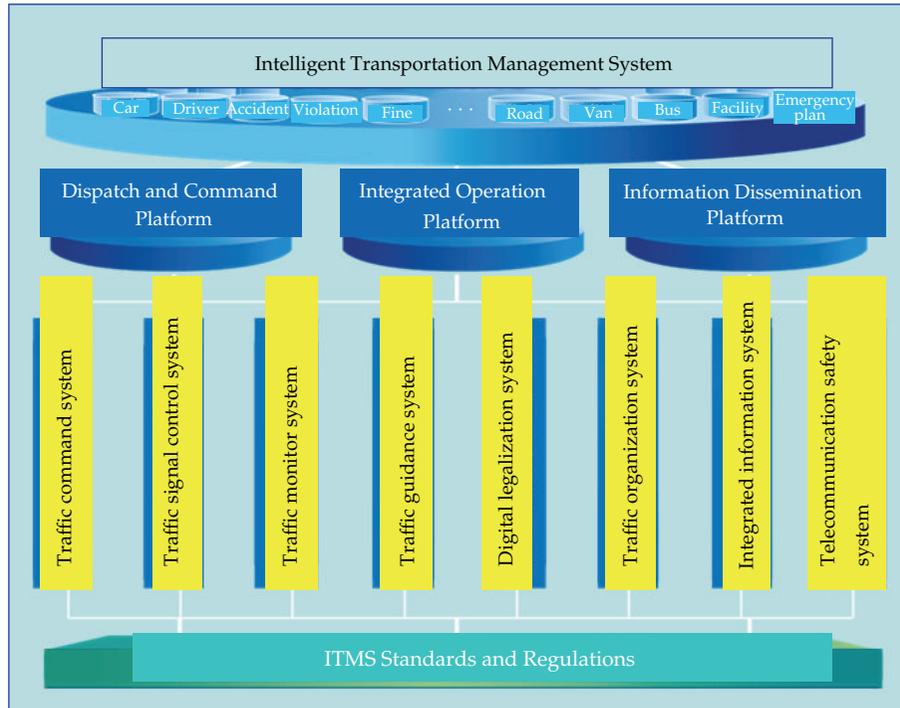


Figure 1: Architecture of ITMS in Beijing.

3 platforms, including the Traffic Management Data Center, the Dispatch and Command Platform, the Integrated Operation Platform, and the Information Dissemination Platform (shown in Figure 1). Based on these center and platforms, 8 related informationized support systems and 99 application subsystems are integrated and developed to mainly function on command and dispatch, traffic signal control, comprehensive supervision, information-provided service, and so forth. With its deployment within the 5th Ring Road in Beijing, the traffic congestion has been alleviated effectively, and some benefits have appeared, such as enhancing traffic capacity, improving service level and road safety, increasing the response to traffic accidents, and reducing environmental pollution. Therefore, more attention has been attracted to study the impacts of ITMS application in Beijing.

1.2. Related Works

The analysis of ITS impact has been a subject of interest for many years; however, the literature mainly emphasized on evaluating ITS benefits qualitatively or quantitatively. On evaluation methods and procedures, in [1], the authors proposed that physical logical structure, market cost, and information interaction structure should be established according to subsystems of ITS, so as to explain the relationship between ITS and estimate socio-economic benefits. Haynes and Li [2] introduced the Probabilistic Multidimensional Scaling Algorithm to evaluate the questionnaires and deploy the subsystems of ITS. The method of Multicriteria Analysis was also applied in ITS impacts evaluation [3]. Both cost-benefit analysis and data envelopment analysis (DEA) [4], one type of multicriteria appraisals,

were applied to analyze the socioeconomic impact of convoy driving systems. A computable general equilibrium (CGE) model was put forward to analyze the impact of ITS on Japan's economy [5]. In addition, a simulation was used to assess the impacts of ITS on the traffic system [6, 7]. Dotoli et al. [8] provided a modeling approach by Petri nets to analyze the impact of Information and Communication Technologies (ICT) on real-time management and operation, as well as the impact on the infrastructures of intermodal transportation systems. Wang et al. [9, 10] brought forward a methodology for model-based digital driving dependability analysis in ITS, which represents an activity in the direction of safety assessment to ITS and gives a new idea how to model digital driving reliability and safety. Recently, Cantarella [11] presented a day-to-day dynamic model, whose application shows that this proposed approach can be applied to model the effect of ITS based both on the value of user surplus and its stability over time.

On ITS impacts and benefits, existing researches such as [12–16] mainly focused on the quantitative evaluation of ITS socioeconomic benefits, and some focused on safety, for example, [7, 17, 18]. Farooq et al. [19] pointed out that ITS affect not only the transportation industry, but also other industries with a quantitative economic analysis through a Leontief's Input-Output (I-O) model [20]. The comprehensive effects, including the environmental and safety benefits, have also been considered in related researches, for example, [21–23].

These efforts have been limited in what ITS effects are and how to evaluate them, but did not refer to how ITS impact the traffic system and why they make effects. Although Newman-Askins et al. [24] pointed out that there was presently little understanding of the causal relationships between ITS projects and their impacts and often it may not be appropriate to transfer results in space and time, they only provided a state-of-practice summary of ITS evaluation methods and impact measurement efforts, without quantitative analysis of the relationship and impact mechanism.

1.3. Motivation and Contribution

To evaluate the influences of ITMS in Beijing comprehensively, it's necessary to identify the impact mechanism of ITMS on the road traffic system, and then define the evaluation content and evaluation system of ITMS effects. The paper is organized as follows. In the next section, we discuss the correlative factors between ITMS and the road traffic system and present 3 positive feedback chains by systemic dynamics theory. In Section 3, according to correlative factors, the intrinsic impact mechanism of ITMS on the road traffic system is studied based on 3 positive feedback chains, respectively. Firstly, the ITMS impact mechanism on travel velocity is illuminated through the Fundamental Diagram (FD) theory. Then with the entropy theory of traffic system, the ITMS impacts on the disequilibrium of road network load and traffic accidents are investigated. Consequently, in terms of the impact mechanism, the effects and benefits brought by ITMS are identified and related evaluation standards are defined in Section 4, which will lay the pretheoretical basis for comprehensive evaluation of ITMS system benefits. Conclusions and an outlook are given in Section 5.

2. Correlative Factors between ITMS and Road Traffic System

It is known that with socioeconomic acceleration of urban development, the vehicle maintenance number increases rapidly and various types of traffic demands expand greatly. Those factors have resulted in the increase of road traffic flow and decrease in traveling speed,

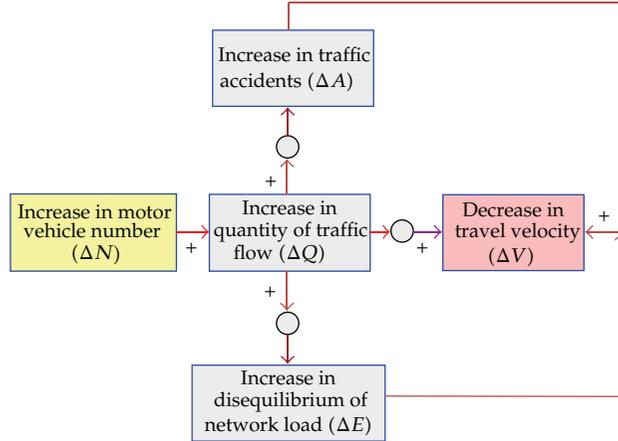


Figure 2: Correlation framework of traffic systematic factors.

and then boosted the occurrence of traffic accidents and enhanced the disequilibrium of road network load. Conversely, the increase of traffic accidents and disequilibrium of network load further impact the decline of traveling speed. We take these factors into consideration and analyze their interactions by the theory of system dynamics. Then we can obtain 3 positive feedback chains as follows:

- (1) the growth in motor vehicle number (ΔN) \rightarrow the increase in quantity of traffic flow (ΔQ) \rightarrow the decrease in travel velocity (ΔV),
- (2) the growth in motor vehicle number (ΔN) \rightarrow the increase in quantity of traffic flow (ΔQ) \rightarrow the increase in disequilibrium of network load (ΔE) \rightarrow the decrease in travel velocity (ΔV),
- (3) the growth in motor vehicle number (ΔN) \rightarrow the increase in quantity of traffic flow (ΔQ) \rightarrow the increase in traffic accidents (ΔA) \rightarrow the decrease in travel velocity (ΔV).

The correlation and interaction between above factors in detail is shown in Figure 2.

Based on above analysis, we introduce the ITMS system into this correlation framework to study its influence. There is no doubt that the deployment of ITMS will exert significant influences on controlling the traveling speed, reducing traffic accidents, and decreasing the disequilibrium of road network load. In all 3 positive feedback chains, ITMS plays a role of negative feedback factor as shown in Figure 3.

According to the ITMS impact mechanism described in Figure 3, we will collect the urban road traffic-related data of Beijing to analyze these 3 feedback chains in the next section.

3. Impact Mechanisms of ITMS on Road Traffic System

3.1. Positive Feedback Chain: $\Delta N \rightarrow \Delta Q \rightarrow \Delta V$

We introduce the ITMS impact into the feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta V$, as shown in Figure 4. As a negative feedback factor, we can find that ITMS plays a role to restrain the decrease of traveling speed.

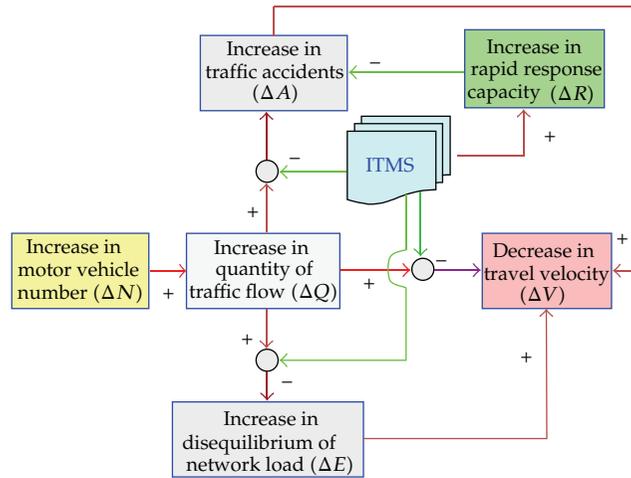


Figure 3: Correlation framework of traffic systematic factors with introducing ITMS as a negative feedback factor.

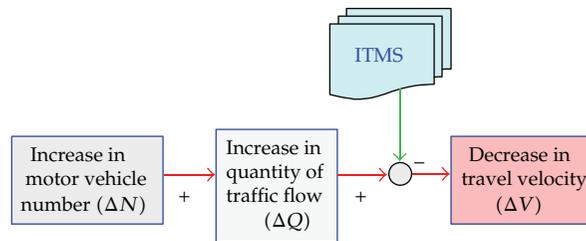


Figure 4: The positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta V$ with the ITMS impact.

In order to study further the impact of ITMS on the urban traffic system, we will analyze the changes in traffic flow and speed against the growth of motor vehicles.

According to Figure 5, it is found that although the motor vehicle number increases rapidly from the year of 2000 to 2010 in Beijing, the growth rate of traffic flow does not maintain consistent with the rate of motor vehicle increase correspondingly. In particular, we can observe that since 2004 when ITMS was begun to deploy in large scale, in spite of the rapid increase in vehicle number, the growth rate of traffic flow was slowing gradually, and there was an obvious decrease in 2008 when Beijing held the Olympic Games. In addition, the decline rate of ring roads' average velocity did not keep pace with the increase rate of motor vehicle. Since 2004, the decline rate of average velocity was also slowing down. Although in 2007 the average velocity of ring roads dropped to the minimum in recent years, in 2008, it began to rebound and was increasing with ITMS being put into practice in a row.

To explain this phenomenon further, based on the traffic data collected from ring roads of Beijing, we use the FD theory to describe the relationship between travel velocity and traffic flow (shown in Figure 6). According to the forecast by the FD theoretical curve, once traffic flow quantity reaches the saturation value, there will be a sudden drop in average velocity; however, in terms of the real data in Beijing's ring roads, we find that after traffic flow quantity reaches the saturation value of 1,800 veh/h, there is no sudden drop in average

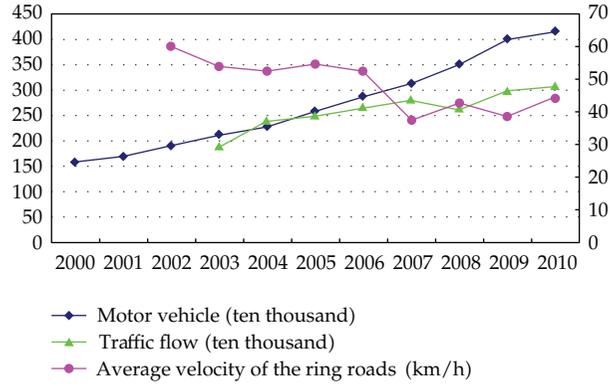


Figure 5: The development trends of motor vehicle maintenance number of Beijing, average traffic flow, and average travel velocity of the 2nd, 3rd, and 4th Ring Roads in Beijing from the year of 2000 to 2010.

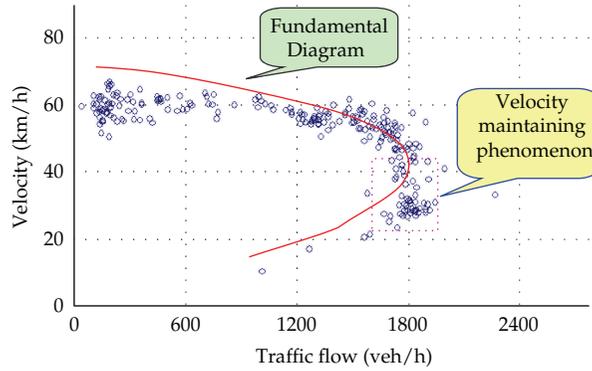


Figure 6: Microcosmic data (denoted by “o”) of average traffic flow and velocity of the 2nd, 3rd, and 4th Ring Roads in Beijing.

velocity, but there is maintaining in a certain range, which we call it *velocity maintaining phenomenon*.

Likewise, we analyze the relationship between average traffic flow and vehicle density (shown in Figure 7). We also find that when the density of vehicles reaches 100 veh/km saturation value, the real traffic flow in Beijing does not decline as rapidly as what is predicted by the FD theory, but is maintaining in a certain range, which we call it *flow rate maintaining phenomenon*.

Based on above empirical analysis of Figures 4, 5, and 6, abstractly, we can consider that ITMS, as a negative feedback factor, has a damping function on the coupling relationship of positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta V$. To describe this damping function, we put forward the concepts of *speed maintaining capability* and *flow rate maintaining capability*. The definitions are identified as follows.

3.1.1. Speed Maintaining Capability C_v

Speed maintaining capability C_v means below the critical speed of saturated traffic flow, a rate of traffic flow samples that are with velocity maintaining phenomenon on the overall

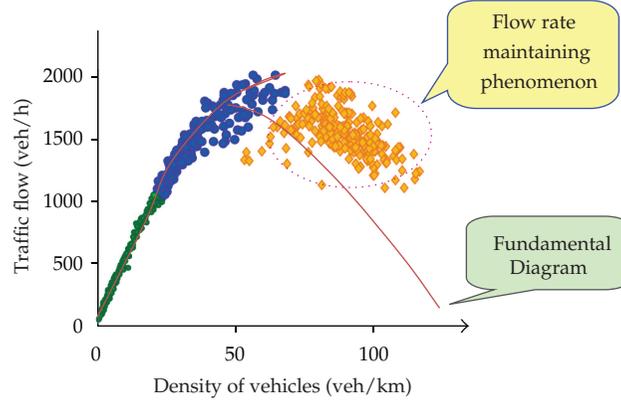


Figure 7: Microcosmic data (denoted by “o”) of average traffic flow and vehicle density of the 2nd, 3rd, and 4th Ring Roads in Beijing.

population. It indicates how much traffic flow can maintain in a certain velocity range without a sudden drop after traffic flow quantity reaches the saturation value. C_v is demonstrated as

$$C_v = \frac{n(v, q)|_{v \in [v_c - \Delta v, v_c]}}{N(v, q)|_{v \leq v_c}} \times 100\%, \quad (3.1)$$

where $v_c : v|_{q=q_{\max}}$ is the critical speed when traffic flow quantity $q = q_{\max}$. Here, v_c, q_{\max} can be defined by a approach presented by one of our authors—Guan and He [25]; Δv demonstrates the increment of velocity; $n(v, q)$ is the sample size of traffic flow when $v \in [v_c - \Delta v, v_c]$; $N(v, q)$ is the overall population when $v \leq v_c$.

3.1.2. Flow Rate Maintaining Capability C_q

Flow rate maintaining capability C_q is a rate of the traffic flow samples with flow rate maintaining phenomenon on the overall population over vehicle critical density. It reflects a degree which the quantity of traffic flow can maintain in a certain flow rate range without a sudden drop since the density of vehicles is larger than the saturation value. C_q is demonstrated as

$$C_q = \frac{n(q, \rho)|_{\rho \geq \rho_c, q \in [q_{\max} - \Delta q, q_{\max}]}}{N(q, \rho)|_{\rho \geq \rho_c}} \times 100\%, \quad (3.2)$$

where $\rho_c : \rho|_{q=q_{\max}}$ is the critical density when traffic flow quantity $q = q_{\max}$. Here, ρ_c can be defined by [25]; Δq demonstrates the traffic flow increment of volume; $n(q, \rho)$ is the sample size when vehicle density $\rho \geq \rho_c$ and traffic flow quantity $q \in [q_{\max} - \Delta q, q_{\max}]$; $N(q, \rho)$ is the overall population when $\rho \geq \rho_c$.

In the rest of this section, we will calculate C_v and C_q by above equations, based on the traffic flow data of the 2nd, 3rd, and 4th Ring Roads that are collected by Remote

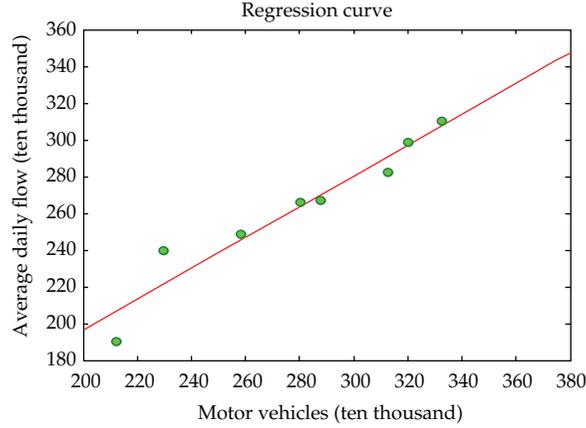


Figure 8: Linearity correlation between motor vehicle number and traffic flow quantity in Beijing.

Traffic Microwave Sensors (RTMS) set up in Beijing, and make an analysis for the computing results. Before computing C_v and C_q , the data outliers have been preprocessed in advance with a method that is presented by Guan and He [26]. Here, the impacts by changes of roads' condition are not considered.

Firstly, through the regression analysis, it is not difficult to find that the two parameters of vehicle number and traffic flow quantity are highly coupled, and the correlation coefficient is 0.96 (shown in Figure 8), which indicates that each vehicle in Beijing has nearly traveled one trip on the 2nd, 3rd, or 4th Ring Road every day on average.

Secondly, according to the phase identification of traffic flow on urban freeways [25] in Beijing, we can define $q_{\max} = 2000$ veh/h and then v_c is 40 km/h. Thus, when $v_c = 40$ km/h, the *speed maintaining capability* C_v of ring roads is calculated as follows:

- (i) when $\Delta v = 5$ km/h, $C_v = 17.0\%$;
- (ii) when $\Delta v = 10$ km/h, $C_v = 45.3\%$;
- (iii) when $\Delta v = 15$ km/h, $C_v = 90.6\%$.

This means, for example, after the flow rate reaches its maximum value ($q_{\max} = 2000$ veh/h), the sample data whose speed is between 25 km/h to 40 km/h account for nearly 90% of total population. It reflects that more than 90% of traffic flow still can travel quite smoothly without a sudden transform to stop status after traffic flow quantity reaches the saturation value.

Thirdly, when $\rho_c = 50$ veh/km ($q_{\max} = 2000$ veh/h), the *traffic flow maintaining capability* C_q of ring roads are as follows:

- (i) when $\Delta q = 500$ veh/km, $C_q = 47.0\%$;
- (ii) when $\Delta q = 800$ veh/km, $C_q = 83.1\%$.

This means that when the density is larger than the critical density ($\rho_c = 50$ veh/km), the sample data whose flow rate is between 1500 veh/h to 2000 veh/h account for 47% of total population, and the sample data whose flow rate is between 1200 veh/h to 2000 veh/h account for 83% of total population. The computing results show that although the density is over the critical value, quite large numbers of traffic flow can maintain in a certain flow rate range without a sudden drop.

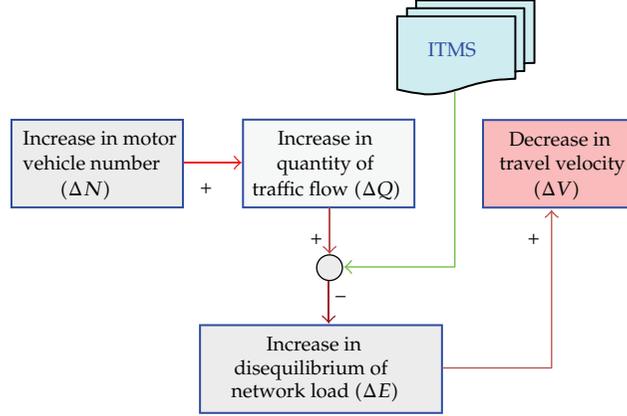


Figure 9: The Positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta E \rightarrow \Delta V$ with the ITMS impact.

Consequently, we can draw a conclusion that ITMS, as a negative feedback factor, has a damping function on the coupling relationship in positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta V$.

3.2. Positive Feedback Chain: $\Delta N \rightarrow \Delta Q \rightarrow \Delta E \rightarrow \Delta V$

In the positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta E \rightarrow \Delta V$, we also introduce the ITMS factor (shown in Figure 9).

Similarly, we need to study the effect and nature of ITMS on the coupling relationship in the positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta E \rightarrow \Delta V$. Here, the *entropy* of the system science theory will be introduced to analyze ITMS impact on this positive feedback chain as follows.

The *entropy* is a state function to indicate the rate of absorbed heat on temperature for material during a reversible process (by Rudolf Clausius in 1865). It is a parameter used to reflect the disorder degree for a system. The larger the entropy is, the more chaotic a system would be. The *entropy ds* of an open system consists of two parts: one is the entropy increase called *entropy production* $d_i s$ that is caused by an irreversible process within the system itself; the other is the entropy exchange called *entropy flow* $d_e s$ that is caused by exchanging the energy and substance between the system and the outside. The total entropy change ds of an opened system is the sum of *entropy production* and *entropy flow*, that is,

$$ds = d_i s + d_e s. \quad (3.3)$$

In the research field of traffic flow, the entropy theory has been introduced to investigate the traffic systematic characteristics and applied in different cases. For example, on modeling research, the traffic system model based on thermodynamics entropy was constructed to reflect the traffic conditions in [27]. Karmakar and Majumder [28] presented the entropy maximization technique to maximize the flow rate at a given continuous traffic stream, so as to illustrate the flow-concentration relationship. Furthermore, Li et al. [29] considered that trip distribution problems can be modeled as entropy maximization models with quadratic cost constraints, then they proposed an entropy maximization model with

chance constraint and proved it is convex. Lu et al. [30] constructed the entropy solutions for the Lighthill-Whitham-Richards (LWR) traffic flow model with a piecewise quadratic flow-density relationship, which may be used for predicting traffic or as a diagnosing tool to test the performance of numerical schemes. One of latest researches on entropy is that an entropy space method is based on a 3D-space built on a flow-packet level by Velarde-Alvarado et al. [31]. And they modeled the network traffic by using Gaussian Mixtures and Extreme Generalized Distributions. By integrating this model in an Anomaly-Based Intrusion Detection System, anomalous behavior traffic can be detected easily and early. Ngoduy and Maher [32] put forward an effort to find global optimal parameters of a second order macroscopic traffic model using a cross entropy method. On entropy-based applications on traffic system, Montemayor-Aldrete et al. [33] introduced a new concept of the production rate of entropy due to traffic flow. With the use of such a concept, the percentage of increase in the fuel consumption rate due to velocity fluctuations on the traffic flow can be easily determined. Tapiador et al. [34] provided a new measure of Intermodal Entropy to help to characterize high speed train stations (HSTS) and improve their intermodal performance. This variable can also be integrated into models of regional accessibility to take into account the intermodality of HSTS within a single comprehensive estimate. Murat et al. [35] used Shannon Entropy Approach to deal with the determination of black spots' traffic safety levels. In addition, the entropy was applied in traffic assignment problems in [29, 36, 37] and the cross entropy method was used to optimize signal [38]. Chi et al. [39] assumed that each cell of the transportation subnetwork O-D flow table contains an elastic demand function and then proposed a combined maximum entropy-least squares (ME-LS) estimator, by which O-D flows are distributed over the subnetwork so as to maximize the trip distribution entropy, while demand function parameters are estimated for achieving the least sum of squared estimation errors. Although the above entropy-based methods and models are different in different studies, they all show that the essence of entropy could be used to reflect a system's characteristics as an effective approach.

Here, based on the dissipative structure theory, we assess the disequilibrium of road network load in terms of the changes in order degree of an open traffic system by value of system entropy [40].

- (1) If $ds > 0$, it shows that traffic system entropy has an increase and the order degree of system will decrease, so it indicates the disequilibrium of road network load will boost;
- (2) If $ds < 0$, it reflects that the external environment has provided negative *entropy flow* to the system, and if $d_{es} < -d_{is}$, the system entropy has a decrease and order degree will increase, so it means the disequilibrium of road network load will decrease;
- (3) If $ds = 0$, it indicates that the order degree of system, that is, the disequilibrium of road network load, is essentially invariable.

In order to calculate the value of entropy in a traffic system, the following will introduce the definition of *information*. *Information* is a measure to the degree of uncertainty for a system. For a given random variable X , the probability of the event x_i is $p(x_i)$, then the *self-information* $I(x_i)$ of the event x_i is defined as, [41],

$$I(x_i) = -C \log_{10} p(x_i), \quad x_i \in X, \quad (3.4)$$

where C is a constant.

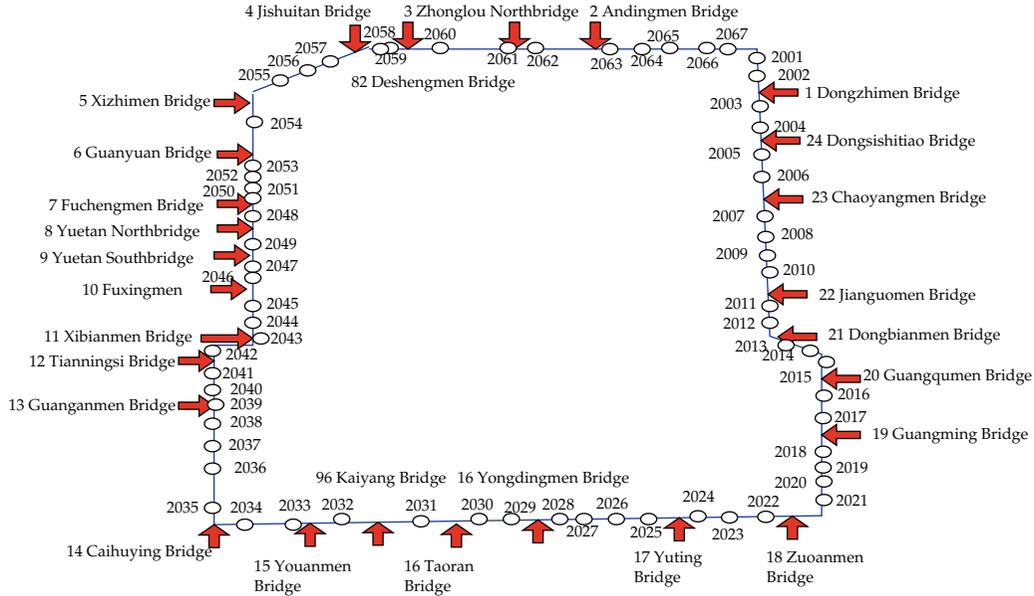


Figure 10: The 2nd Ring Road in Beijing marked with 24 sections.

In 1948, Claude Elwood Shannon introduced the *entropy* into the concept of *information*. And the *entropy* $H(x)$ of a discrete random variable X is identified as

$$H(x) = -C \sum_{i=1}^n p(x_i) \log_{10} p(x_i), \tag{3.5}$$

where C is a constant.

It can be found that the *entropy* is a statistical probability-weighted average of *self-information* in terms of (3.4) and (3.5). The *self-information* $I(x_i)$ and *entropy* $H(x)$ are the functions of probability and unrelated to the value of the variables.

According to the above definitions, we use the concept of *traffic entropy* to estimate the disequilibrium of road network load. The *traffic entropy* $E(t)$ is demonstrated as, [42],

$$E(t) = -C \sum_{i=1}^n k_i \cdot p_i \cdot \log_{10} p_i, \tag{3.6}$$

where $E(t)$ is the entropy of a road traffic system, and C is a constant; k_i is the weight of the road section i , $\sum_{i=1}^n k_i = 1$; $p_i = \{\Delta \lambda_i / \sum_{i=1}^n \Delta \lambda_i, i = 1, 2, 3, \dots, n\}$ is distribution of congestion deviation 9, $\sum_{i=1}^n p_i = 1$; $\Delta \lambda_i = \max\{(\lambda_i - 0.8), 0\}$ is congestion deviation of the road section i ; λ_i denotes the average congestion degree, and it is a rate of traffic flow on capacity in section i ; and $\lambda = 0.8$ is threshold of congestion.

Then we will illustrate the meaning of *traffic entropy* by calculating entropy of the 2nd Ring Road of Beijing in 2004 and 2008, respectively. The 2nd Ring Road shown in Figure 10 is marked with 24 sections, that is, $i = 1, 2, \dots, 24$.

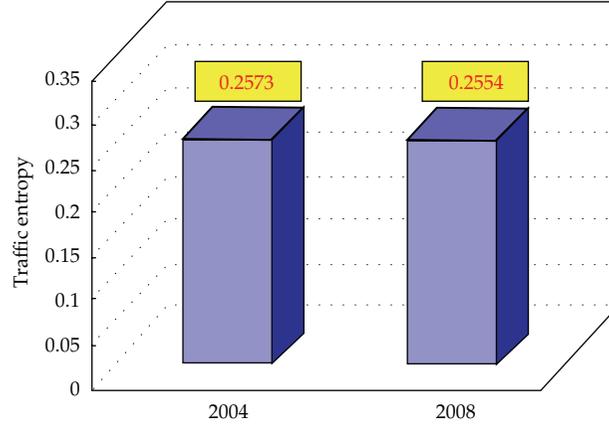


Figure 11: Traffic entropies of the 2nd Ring Road in 2004 and 2008 in Beijing.

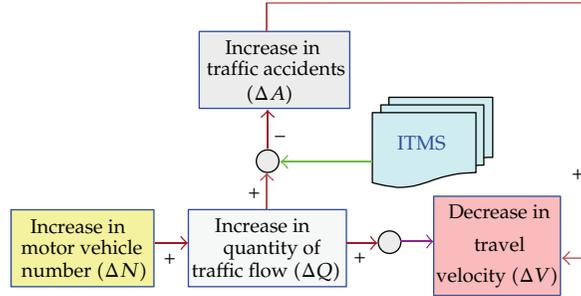


Figure 12: The positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta A \rightarrow \Delta V$ with the ITMS impact.

Based on the empirical data, we estimate the traffic entropies of the 2nd Ring Road in 2004 and 2008 by (3.5). The comparison of these two entropies is shown in Figure 11.

Figure 11 shows that the traffic entropy of 2008 does not increase, but decrease slightly comparing with that of 2004, which means that the order degree has increased and the disequilibrium of road network load decreased despite the rapid growth in vehicle number. Therefore, we can get a conclusion that ITMS, as a negative feedback factor, has a damping function on the coupling relationship of positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta E \rightarrow \Delta V$.

3.3. Positive Feedback Chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta A \rightarrow \Delta V$

In the same way, we introduce the ITMS impact into the positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta A \rightarrow \Delta V$, which is shown in Figure 12.

In terms of the statistic date, the number of traffic accidents in 2008 has decreased from 5,425 to 1,960 comparing with that of 2004 (shown in Figure 13), with the decreased accidents accounting for about 62.03% of 2004; while the number of vehicles is growing increasingly, the total length of road has extended from 4,064 km to 6,186 km and the total area of road from 64.17 million m^2 to 89.40 million m^2 . It indicates that the safety of urban traffic system has been improved with ITMS deployment and application. So we can draw a conclusion

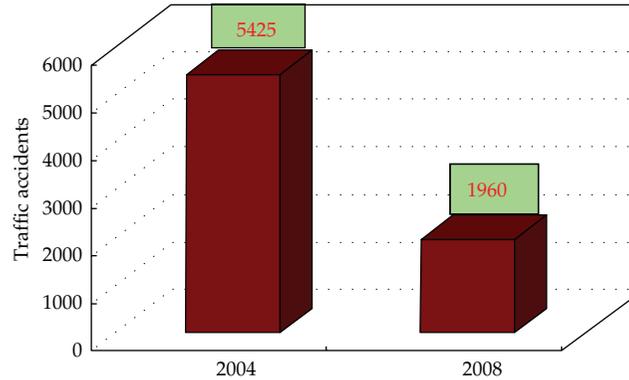


Figure 13: Comparison of traffic accidents number between 2004 and 2008 in Beijing.

that ITMS has a damping function on the coupling relationship of positive feedback chain $\Delta N \rightarrow \Delta Q \rightarrow \Delta A \rightarrow \Delta V$ as a negative feedback factor.

4. Benefits and Evaluation Standard of ITMS on Road Traffic System

Based on above analysis, it can get a conclusion that with the deployment of ITMS in Beijing, the negative influence on traveling speed, road traffic safety, and the disequilibrium of network load aroused by the increase of traffic flow have been weakened efficiently. ITMS has effects on improving road traffic efficiency, alleviating traffic congestion, and enhancing road service capacity. According to the impact mechanism, the benefits brought by ITMS can be concluded in 4 aspects: socioeconomic benefit, energy resource and environmental benefits, traffic safety, and management efficiency.

For each benefit aspect, we identify standards for evaluating benefits of ITMS. The evaluation standards of socioeconomic benefit can be mostly derived from the increase of travel velocity, mainly including (1) reduction in vehicles' operational cost, (2) reducing traveling time, and (3) investment reduction on land resource and transportation infrastructure (shown in Figure 14). And the standards of energy resources and environmental benefits consist of (1) decrease in traffic noise, (2) decrease in vehicle tailpipe emission, (3) greenhouse gas emission, (4) reduction of energy resource, and (5) improvement on consumption structure of energy resource, which are also brought mainly because of the increase in travel velocity (shown in Figure 14).

For evaluation standards of traffic safety benefits, because of the decrease in traffic accidents and improvement on vehicle safety, they can be denoted by (1) reduction in vehicle loss, (2) loss of vehicle stolen or robbed, (3) decrease in loss of personal casualty, and (4) loss of social public institution service (shown in Figure 15). The management efficiency for road administrations can be indicated by the increase in response time for alarm (shown in Figure 15).

5. Conclusion

Summing up, in this paper, in order to investigate the ITMS impacts on road traffic system, we have introduced 3 positive feedback chains to analyze the correlative factors between

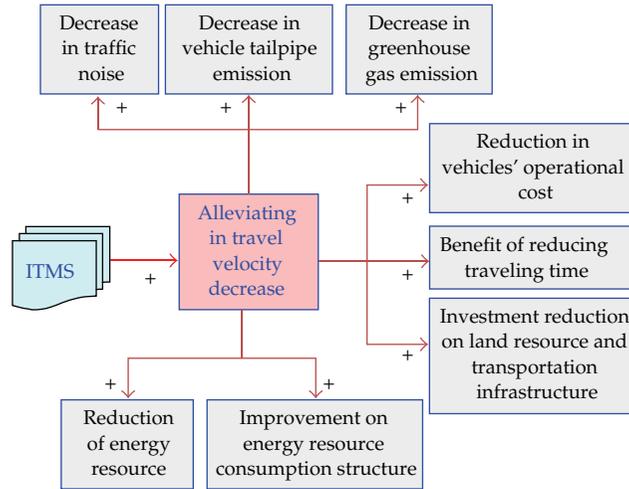


Figure 14: Benefits of social economy, energy resource and environment, and related evaluation standards.

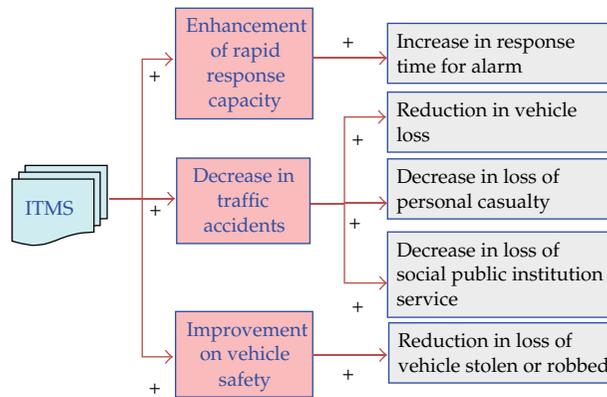


Figure 15: Benefits of traffic safety and management efficiency and related evaluation standards.

ITMS and road traffic system. The results show that ITMS has damping functions on the coupling relationship of 3 positive feedback chains $\Delta N \rightarrow \Delta Q \rightarrow \Delta V$, $\Delta N \rightarrow \Delta Q \rightarrow \Delta E \rightarrow \Delta V$, and $\Delta N \rightarrow \Delta Q \rightarrow \Delta A \rightarrow \Delta V$ as a negative feedback factor. This indicates that with its deployment in Beijing, ITMS has impacted significantly the improvement on road traffic efficiency, alleviation of traffic congestion, and enhancement of road service capacity. Accordingly, in terms of this impact mechanism, the benefits of ITMS deployment have been presented and related evaluation standards are identified, which provides a comprehensive evaluation framework for subsystems of ITS. Such analyses would provide richer information for ITS benefits evaluation. Indeed, some policy factors, such as traffic regulation and limitation of vehicle use during period of 2008 Olympic Games, are not taken in account to analyze the impact mechanism. But our paper present an approach to study the ITS impacts from a new view point. Given more data, future work would focus on considering the above policy factors and quantifying the impact benefits.

Acknowledgments

This research is supported by the Major State Basic Research Development Program of China (973 Program) (Grant no. 2012CB725403-5), the Key Program of National Natural Science Foundation of China (Grant no. 71131001-2), and the Fundamental Research Funds for the Central Universities of China (Approval no. 2011JBM062).

References

- [1] H. Lo, M. Hickman, and S. Weissenberger, "Structured approach for ITS architecture representation and evaluation," in *Proceedings of the 6th Vehicle Navigation and Information Systems Conference*, pp. 442–449, August 1995.
- [2] K. E. Haynes and M. Li, "Analytical alternatives in intelligent transportation system (ITS) evaluation," *Research in Transportation Economics*, vol. 8, pp. 127–149, 2004.
- [3] K. D. Brucker, A. Verbeke, and C. Macharis, "The application of multicriteria analysis to the evaluation of intelligent transport systems," *Research in Transportation Economics*, vol. 8, pp. 151–179, 2004.
- [4] Z.-C. Juan, J.-P. Wu, and M. McDonald, "Socio-economic impact assessment of intelligent transport systems," *Tsinghua Science and Technology*, vol. 11, no. 3, pp. 339–350, 2006.
- [5] T. Kawakami, P. Tiwari, and M. Doi, "Assessing impact of ITS on Japan's economy using a computable model," *Research in Transportation Economics*, vol. 8, pp. 525–547, 2004.
- [6] H. Liu, L. Chu, and W. Recker, "Performance evaluation of ITS strategies using microscopic simulation," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (ITSC '04)*, pp. 917–921, Washington, DC, USA, October 2004.
- [7] M. Vanderschuren, "Safety improvements through Intelligent Transport Systems: a South African case study based on microscopic simulation modelling," *Accident Analysis and Prevention*, vol. 40, no. 2, pp. 807–817, 2008.
- [8] M. Dotoli, M. P. Fantì, A. M. Mangini, G. Stecco, and W. Ukovich, "The impact of ICT on intermodal transportation systems: a modelling approach by Petri nets," *Control Engineering Practice*, vol. 18, no. 8, pp. 893–903, 2010.
- [9] W.-H. Wang, H.-W. Guo, H. Bubb, and K. Ikeuchi, "Numerical simulation and analysis procedure for model-based digital driving dependability in intelligent transport system," *KSCE Journal of Civil Engineering*, vol. 15, no. 5, pp. 891–898, 2011.
- [10] W.-H. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, China Communications Press, Beijing, China, 2012.
- [11] G. E. Cantarella, "Day-to-day dynamic models for intelligent transportation systems design and appraisal," *Transportation Research Part C*. In press.
- [12] L. K. Pekka and L. H. Jukka, "Profitability evaluation of intelligent transport system investments," *Journal of Transportation Engineering*, vol. 128, no. 3, pp. 276–286, 2002.
- [13] Y. J. Nakanishi and J. C. Falcocchio, "Performance assessment of intelligent transportation systems using data envelopment analysis," *Research in Transportation Economics*, vol. 8, pp. 181–197, 2004.
- [14] T. Moore and R. Pozdena, "Framework for an economic evaluation of transportation investments," *Research in Transportation Economics*, vol. 8, pp. 17–45, 2004.
- [15] A. Stevens, "The application and limitations of cost-benefit assessment (CBA) for intelligent transport systems," *Research in Transportation Economics*, vol. 8, pp. 91–111, 2004.
- [16] W. Chris and M. Mahmoud, "Transportation benefit-cost analysis: lessons from CAL-B/C," in *Proceedings of the 87th Annual Meeting of Transportation Research Board*, pp. 79–87, Washington, DC, USA, 2004.
- [17] T. Vaa, M. Penttinen, and I. Spyropoulou, "Intelligent transport systems and effects on road traffic accidents: state of the art," *IET Intelligent Transport Systems*, vol. 1, no. 2, pp. 81–88, 2007.
- [18] R. Kulmala, "Ex-ante assessment of the safety effects of intelligent transport systems," *Accident Analysis and Prevention*, vol. 42, no. 4, pp. 1359–1369, 2010.
- [19] U. Farooq, M. A. Siddiqui, L. Gao, and J. L. Hardy, "Intelligent transportation systems: an impact analysis for michigan," *Journal of Advanced Transportation*, vol. 46, pp. 12–25, 2012.
- [20] W. Leontief, "Quantitative input and output relations in the econometric system of the United States," *The Review of Economic Statistics*, vol. 18, pp. 105–125, 1936.
- [21] B. J. Kanninen, "Intelligent transportation systems: an economic and environmental policy assessment," *Transportation Research A*, vol. 30, no. 1, pp. 1–10, 1996.

- [22] J.-C. Thill, G. Rogova, and J. Yan, "Evaluating benefits and costs of intelligent transportation systems elements from a planning perspective," *Research in Transportation Economics*, vol. 8, pp. 571–603, 2004.
- [23] F. Shi, J. Lu, and W. Wang, "Method of benefits evaluation for urban traffic command center and its application," *Computer and Communications*, vol. 22, no. 2, pp. 43–46, 2004.
- [24] R. Newman-Askins, L. Ferreira, and J. Bunker, "Intelligent transport systems evaluation: from theory to practice," in *Proceedings of the 21st ARRB and 11th REAAA Conference, Transport Our Highway to a Sustainable Future*, pp. 1011–1026, May 2003.
- [25] W. Guan and S.-Y. He, "Statistical features and phase identification of traffic flow on urban freeway," *Journal of Transportation Systems Engineering and Information Technology*, vol. 7, no. 5, pp. 42–50, 2007.
- [26] W. Guan and S.-Y. He, "Statistical features of traffic flow on urban freeways," *Physica A*, vol. 387, no. 4, pp. 944–954, 2008.
- [27] C.-Y. Sheng, Y.-L. Pei, and L.-N. Wu, "Traffic entropy characteristics at on-ramp junctions on urban freeway," *Journal of Jilin University*, vol. 37, no. 6, pp. 1273–1277, 2007.
- [28] K. Karmakar and S. K. Majumder, "Maximum entropy approach in a traffic stream," *Applied Mathematics and Computation*, vol. 195, no. 1, pp. 61–65, 2008.
- [29] X. Li, Z.-F. Qin, L.-X. Yang, and K.-P. Li, "Entropy maximization model for the trip distribution problem with fuzzy and random parameters," *Journal of Computational and Applied Mathematics*, vol. 235, no. 8, pp. 1906–1913, 2011.
- [30] Y.-D. Lu, S. Wong, M.-P. Zhang, C.-W. Shu, and W.-Q. Chen, "Explicit construction of entropy solutions for the lighthill-whitham-richards traffic flow model with a piecewise quadratic flow—density relationship," *Transportation Research B*, vol. 42, pp. 355–372, 2008.
- [31] P. Velarde-Alvarado, A. F. Martínez-Herrera, and A. Iriarte-Solis, "Using entropy spaces and mixtures of gaussian distributions to characterize traffic anomalies," *Procedia Technology*, vol. 3, pp. 97–108, 2012.
- [32] D. Ngoduy and M. J. Maher, "Calibration of second order traffic models using continuous cross entropy method," *Transportation Research C*, vol. 24, pp. 102–121, 2012.
- [33] J. A. Montemayor-Aldrete, P. Ugalde-Vélez, M. D. Castillo-Mussot, C. A. Vázquez-Villanueva, G. J. Vázquez, and A. Mendoza-Allende, "Steady state traffic flow, entropy production rate, temporal fluctuations and fuel consumption," *Physica A*, vol. 361, no. 2, pp. 630–642, 2006.
- [34] F. J. Tapiador, K. Burckhart, and J. Martf-Henneberg, "Characterizing European high speed train stations using intermodal time and entropy metrics," *Transportation Research A*, vol. 43, no. 2, pp. 197–208, 2009.
- [35] Y. S. Murat, "An entropy (shannon) based traffic safety level determination approach for black spots," *Procedia Social and Behavioral Sciences*, vol. 20, pp. 786–795, 2011.
- [36] H. Bar-Gera, "Traffic assignment by paired alternative segments," *Transportation Research B*, vol. 44, no. 8-9, pp. 1022–1046, 2010.
- [37] S. Lu and Y. Nie, "Stability of user-equilibrium route flow solutions for the traffic assignment problem," *Transportation Research B*, vol. 44, no. 4, pp. 609–617, 2010.
- [38] M. Maher, R.-h. Liu, and D. Ngoduy, "Signal optimisation using the cross entropy method," *Transportation Research Part C: Emerging Technologies*. In press.
- [39] X. Chi, K. M. Kockelman, and T. Waller, "A maximum entropy-least squares estimator for elastic origindestination trip matrix estimation," *Procedia Social and Behavioral Sciences*, vol. 17, pp. 189–212, 2011.
- [40] W.-T. Fan, W. Jia, and Y.-M. Ding, "A possible approach to the framework of the fundamental theory of system science: part ten," *System Engineering Theory & Practice*, vol. 8, pp. 19–20, 2002.
- [41] Z.-X. Ye, *Basic Information Theory*, Higher Education Press, Beijing, China, 2003.
- [42] W. Nie, C.-F. Shao, and L.-Y. Yang, "Study on the application of dissipative structure and entropy change theory in regional transportation system," *Journal of Highway and Transportation Research and Development*, vol. 23, no. 10, pp. 1–10, 2006.

Research Article

Fast Pedestrian Recognition Based on Multisensor Fusion

**Hongyu Hu,^{1,2} Zhaowei Qu,² Zhihui Li,²
Jinhui Hu,² and Fulu Wei²**

¹ State Key Laboratory of Automobile Dynamic Simulation, Jilin University,
Changchun 130022, China

² College of Transportation, Jilin University, Changchun 130022, China

Correspondence should be addressed to Zhihui Li, lizhih@jlu.edu.cn

Received 11 September 2012; Revised 10 November 2012; Accepted 21 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Hongyu Hu 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 fast pedestrian recognition algorithm based on multisensor fusion is presented in this paper. Firstly, potential pedestrian locations are estimated by laser radar scanning in the world coordinates, and then their corresponding candidate regions in the image are located by camera calibration and the perspective mapping model. For avoiding time consuming in the training and recognition process caused by large numbers of feature vector dimensions, region of interest-based integral histograms of oriented gradients (ROI-IHOG) feature extraction method is proposed later. A support vector machine (SVM) classifier is trained by a novel pedestrian sample dataset which adapt to the urban road environment for online recognition. Finally, we test the validity of the proposed approach with several video sequences from realistic urban road scenarios. Reliable and timewise performances are shown based on our multisensor fusing method.

1. Introduction

Pedestrians are vulnerable participants among all objects involved in the transportation system when crashes happen, especially those in motion under urban road scenarios [1]. In 2009, it was found that in the first global road safety assessment of World Health Organization report, traffic accident is one of the major causes of death and injuries around the world. 41% to 75% of road traffic fatal accidents are involving pedestrians, and the lethal possibility of pedestrians is 4 times compared with that of vehicle occupants. Therefore, pedestrian safety protection should be taken seriously [2].

Active pedestrian safety protection could be able to avoid collision between vehicles and pedestrians essentially, thus it has become the most promising technique to enhance the safety and mobility of pedestrians. For active safety systems, various types of sensors are utilized to detect pedestrians and generate appropriate warnings for drivers or perform autonomous braking in the case of an imminent collision [3]. To detect pedestrian in real time is one of the most crucial part. The current research is mainly focused on the application of visual sensors [4–6], infrared (IR) imaging sensors [7, 8], and radar sensors [9, 10] to aware of pedestrians and obtain their safety state information for realizing active pedestrian protection. Each type of sensors has its advantages and limitations. In order to enhance the advantages and overcome the limitations, multimodality information fusion has become the development trend of pedestrian detection and security warning [3]. Depending on the complementarity information from different sensors, more reliable and robust pedestrian detection results could be obtained by processing multisource and heterogeneous data. In past two decades, reducing accidents involving vulnerable road users with fusion of different kinds of sensors has already been focused on by some research projects, such as APVRU [11], PAROTO [12], PROTECTOR [13, 14], and SAVE-U [15, 16] in European countries. Some considerable and available researches have been conducted by various groups. Scheunert et al. [17] detected range discontinuities utilized by laser scanner and high brightness region in the image by far infrared sensor (FIR). Data fusion based on Kalman filter handled the combination of the outputs from laser scanner and FIR. Szarvas et al. [18] created a range mapping method to identify pedestrian location and scale by laser radar and camera fusion. Neural network was utilized for image-based feature extraction and object classification. Töns et al. [19] combined the radar sensor, IR, and vision sensor for robust pedestrian detection, classification, and tracking. Bertozzi et al. [20] fused stereo vision and IR to obtain disparity data together for pedestrian detection. Vision was used to preliminary detect the presence of pedestrians in a specific region of interest. Results were merged with a set of regions of interest provided by a motion stereo technique. Combining with a laser scanner and a camera, Broggi et al. presented [21] an application for detecting pedestrian appearing just behind occluding obstacles.

Despite some achievements have already been made, complementary advantages of multisensor data fusion are not fully realized. Multimodality data fusion-based pedestrian detection algorithm should be further improved for higher detection accuracy rate and timewise performance boosting, especially in some complicated urban road environment. This paper aims to propose a real time pedestrian recognition algorithm based on laser scanner and vision information fusion. In the first stage, combining with the information of radar scanning, the pedestrian candidate regions in the image are located by space-image perspective mapping model, which could effectively reduce the computational time cost for pedestrian recognition. In the second stage, ROI-IHOG feature extraction method is proposed for further improving the computation efficiency, which could ensure the real time and reliability of online pedestrian recognition effectively.

The reminder of this paper is organized as follows: a brief overview of proposed pedestrian recognition system is presented firstly, followed by the pedestrian candidate region estimation based on laser scanner and vision information fusion, and then we focus on vision-based pedestrian recognition. Finally, we test the validity of the proposed approach with several urban road scenarios and conclude the works.



Figure 1: Multisensor based pedestrian recognition experimental platform.

2. System Overview

2.1. System Architecture

The research of pedestrian recognition is carried out on the multisensor vehicle platform, as shown in Figure 1. This experimental platform is a modified Jetta. It is equipped with a vision sensor, a laser scanner, and two near-infrared illuminators to detect pedestrians in the range of 90° in front of the vehicle.

The architecture of the proposed pedestrian detection system based on multisensor is shown in Figure 2. The system is running on an Intel Core I5 CPU, 2.27 GHZ, RAM 2.0 GB PC. The system includes offline training and online recognition. For offline training, a novel pedestrian dataset adapt the urban road environment is established first, and then the pedestrian classifier is trained by SVM. For online recognition, a Sony SSC-ET185P camera installed on the top front of the experimental vehicle is used to capture continuous 320×240 image. Potential pedestrian candidate regions are identified in the image through the radar data from a SICK LMS211-S14 laser scanner and the perspective mapping model between world coordinates and image coordinates. For each image, all candidate regions are scaled to 64×128 and judged by the classifier trained offline.

2.2. Sensor Selection

The Sony SSC-ET185P camera has been chosen for several reasons. The camera has a high color reproduction and sharp images. It includes a 18x optical zoom and 12x digital high-quality zoom lens with autofocus, so the camera can capture high quality color images during the day. Although the system is now being tested under daylight conditions, two near-infrared illuminators are mounted on both sides of the laser radar in front of the vehicle, which allow the object detection due to a specific illumination for the extension of its application at night.

The laser scanner is a SICK LMS211-S14. The detection capabilities (scanning angle of 90° , minimum angular resolution of 0.5° up to 81.91 m range) are suitable for our goal. The laser scanner only scans a flat data, the ranging principle is a time-of-flight method, and it measures the round trip time of flight to determine the distance by emitting light pulses to the target. It takes 13 ms of once scanning which could be able to meet the needs of real time.

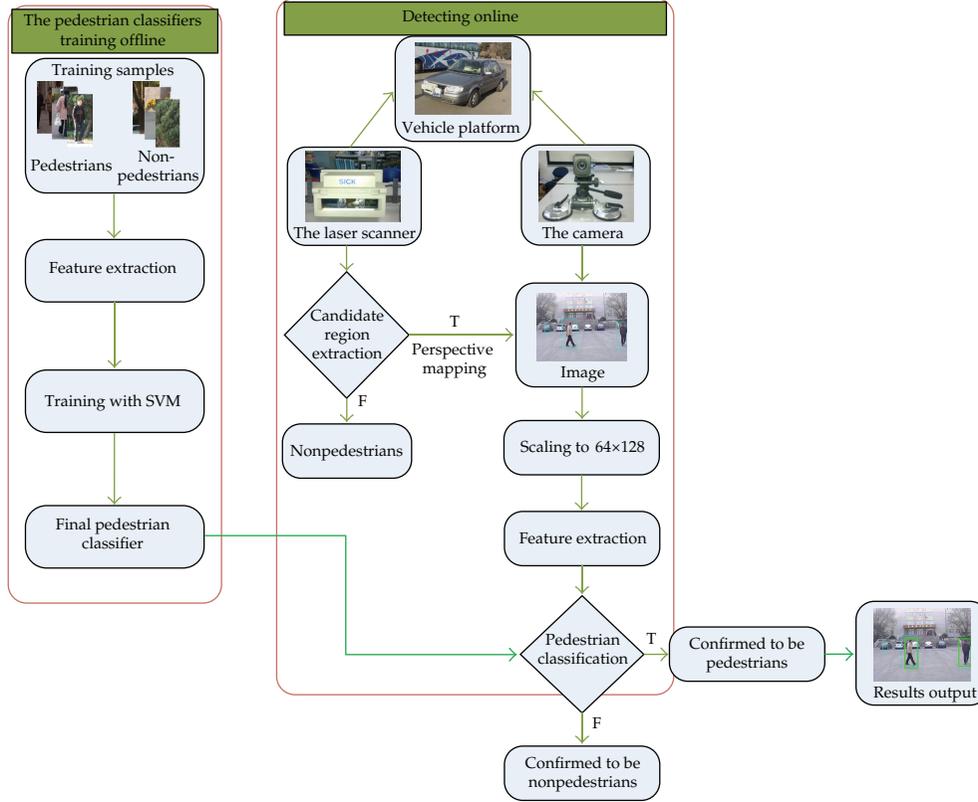


Figure 2: Architecture of the proposed multisensor pedestrian recognition system.

2.3. Vehicle Setup

The laser scanner and two near-infrared illuminators are located in the front bumper in horizontal, as shown in Figure 3(a). The camera is placed at the top front of the vehicle, with the same centerline of the laser scanner, as show in Figure 3(b). The horizontal distance between the camera and the laser scanner is 2.3 m, and the camera height is 1.6 m, which are two key parameters of the camera calibration.

The MINE V-cap 2860 USB is used to connect between the camera and the PC. An RS-422 Industrial serial and MOXA NPort high-speed card provide an easy connection between the laser and PC. Figure 4 shows the hardware integration of the proposed system.

3. Potential Pedestrian Location Estimation

Most current pedestrian detection methods are simply depending on visual sensors that cannot meet the real time application. In our work, we attempt to utilize laser radar sensor to detect obstacle locations for potential pedestrian position estimation in world coordinates, and then we make use of the camera calibration and the space-image perspective mapping model to mark the pedestrian candidate region in the image. Pedestrian recognition algorithm proposed later is performed only for the candidate regions instead of the entire



(a)

(b)

Figure 3: (a) Installation location of the laser scanner and near-infrared illuminators. (b) Installation location of the camera.

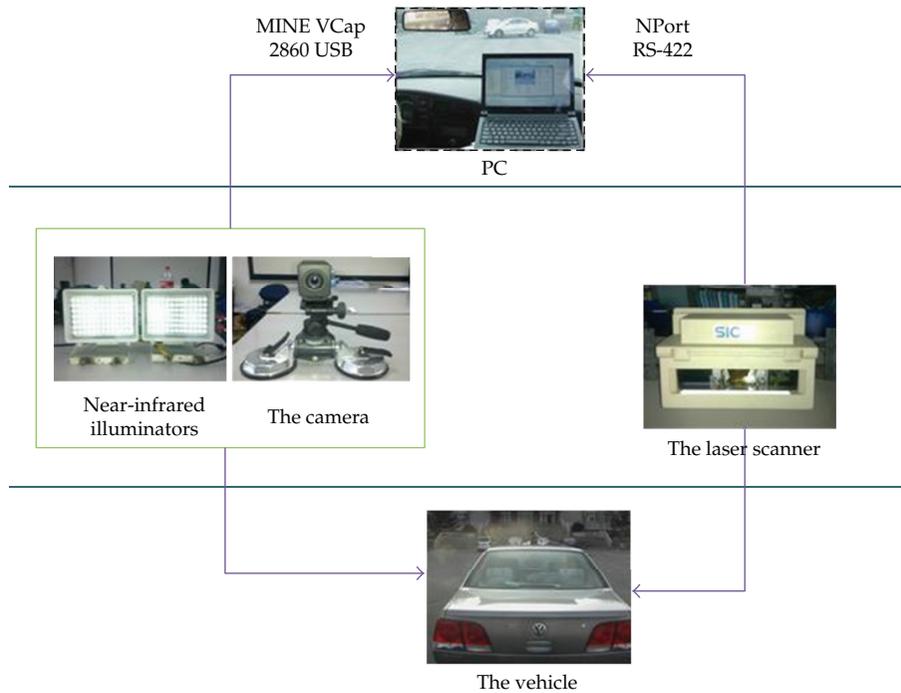


Figure 4: Hardware integration of the platform.

image, which could effectively reduce the computational time cost for a good real time application.

In our experimental platform, a SICK LMS211-S14 laser scanner is utilized. The scanning angle is 90° in front of the host vehicle and the minimum angular resolution of 0.5° (in Figure 5). Thus, we can get 181 data arrays from radar sensor scanning once time. Each data array includes two parameters: the angle and the distance between the obstacle and the host vehicle. A data array could be denoted as $\{(\rho_i, \theta_i) \mid i = 1, 2, \dots, m\}$, where m is the total number of the array, and (ρ_i, θ_i) is the data of the i th array.

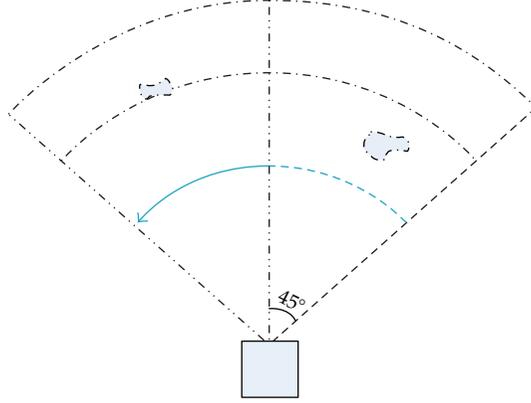


Figure 5: The sketch of radar scanning.

Obviously, a set of laser beams from the same target should have the similar distances and the similar angles. Based on this, a clustering method is applied for 181 data to determine which belong to the same target, which is denoted as

$$D_1 < \sqrt{r_k^2 + r_{k+1}^2 - 2r_k r_{k+1} \cos(\phi)} < D_2, \quad (3.1)$$

where ϕ is the minimum angular resolution of the radar; r_k is the distance of the k th array; D_1 , D_2 are the distance threshold. According to the installation location of the radar, the part of pedestrian's knees would be scan. Taking into account the actual physical characteristics of the pedestrian (legs separated or closed) in spatial, D_1 , D_2 are set as 10 cm and 70 cm, respectively. Then, the potential pedestrian location parameters (the start data, the end data, and the data amount) of each target are recorded. The target distance could be expressed by the average distance of all beams from the target: $\rho_a = (\rho_1 + \rho_2 + \dots + \rho_n) / n$. Its direction could be represented as $\theta_a = (\theta_{j1} + \theta_{j2}) / 2$, where θ_{j1} is first angle value of the target, and θ_{j2} is the last one. Finally, we convert the radar data from polar coordinate to Cartesian coordinate as

$$x = r \sin \theta, \quad y = r \cos \theta, \quad (3.2)$$

where (r, θ) is the data in the polar coordinate; (x, y) is the data in the Cartesian coordinate, which represent the target location in space. The possible pedestrian locations are 2D data in world coordinate. Their corresponding regions in the image are then located by a piecewise camera calibration and the perspective space-image mapping model. This map is projected into the image in order to identify the regions and scale to search for pedestrians in the image. The camera height is 1.6 m, which is a parameter of the camera calibration. We can obtain the space-image mapping model as follows:

$$Y_W \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a_x & u_0 & 0 & 0 \\ 0 & v_0 & -a_z & 1.6a_z \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad \begin{cases} u = \frac{x_w}{y_w} a_x + u_0 \\ v = \frac{1.6 - z_w}{y_w} a_z + v_0, \end{cases} \quad (3.3)$$

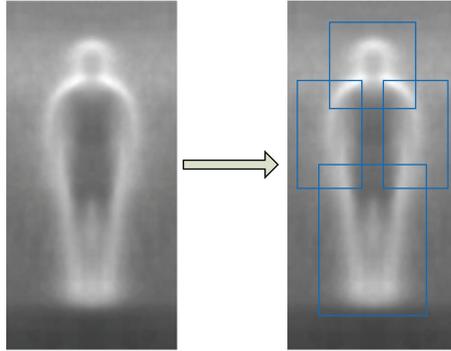


Figure 6: ROI of average samples.

where X_w , Y_w , and Z_w are location parameters in world coordinate; u, v are corresponding parameters in image coordinate. We divided the detection area into four sections which gradually to determine, respectively, the mapping model parameters u, v more accurately by the least square method.

In order to detect pedestrians more accurately and faster, we should determine the detection size of the candidate pedestrian imaging region at different distances in front of the vehicle. We assumed that the pedestrian template is 2 m height and 1 m width (a little larger than real pedestrian). The relationship between the pedestrian's width and height of the imaging region and the pedestrian location in space could be found by the calibration experiment. The potential pedestrian region's width and height in the image could be denoted as $h = 1402y^{-0.97}$, $w = 723.2y^{-0.99}$, where y is vertical distance from the target to the host vehicle.

4. Pedestrian Recognition

4.1. Feature Representation

In 2005, Dalal and Triggs [22] proposed the grids of histogram of oriented gradient (HOG) descriptors for pedestrian detection. Experiment results showed that HOG feature sets significantly outperformed existing feature sets for human detection. However, HOG-based algorithm is too time consuming, especially for multi-scale object detection. The approach should be further optimized because it is not suitable for real time pedestrian safety protection.

In this paper, for fast pedestrian detection, the region of interest (ROI) of a pedestrian sample is found by calculating the average gradient of all positive samples in RSPerson dataset mentioned below. We can find that the gradient features at the head and limbs of pedestrian samples are most obvious. On the other hand, the gradients of the background area in the sample image offer less effect for pedestrian detection which may also disturb the processing performance. Therefore, in order to reduce HOG feature vector dimension of a whole image (3780 dimensions), several important areas are considered as ROI of a selected sample image to calculate the HOG feature. Accordingly, the computation amount of HOG feature is greatly reduced, and pedestrian recognition speed is improved. Through the analysis of average gradient value of pedestrian samples which is shown in Figure 6,

Table 1: Location of ROI in the sample.

ROI	Left-top position of a 64 * 128 sample	Width	Height
Head region	(16, 8)	32	32
Left arm region	(4, 30)	24	40
Right arm region	(36, 30)	24	40
Legs region	(12, 60)	40	56

four regions of interest are identified as ROI: the head region, the leg region, the left arm region, and the right arm region. These regions could be part of the overlaps each other and cover the body's contours basically.

For a color image, gradients of each color channel are calculated. The gradients which have the largest amplitude among three color channel are selected as the the gradient vector of each pixel. Optimal ROI location, width, and height of a sample image is shown in Table 1.

Similar with Dalal's method, for calculating the feature vector of ROI in a detection window, the cell's size is defined as 8×8 pixel, and the block's size is defined as a 2×2 cell. The window's scan step is 8 pixels, the width of a cell. A total of 49 blocks could be extracted in a detection window. For each pixel (x, y) in the image, the gradient vector is denoted as $\Delta g(x, y) = (\partial f(x, y)/\partial x, \partial f(x, y)/\partial y)$. In general, one-dimensional centrosymmetric template operator $[-1, 0, 1]$ is used for calculating the gradient vector:

$$\begin{aligned} g_x(x, y) &= \frac{\partial f(x, y)}{\partial x} = f(x+1, y) - f(x-1, y), \\ g_y(x, y) &= \frac{\partial f(x, y)}{\partial y} = f(x, y+1) - f(x, y-1). \end{aligned} \quad (4.1)$$

Accordingly, the gradient magnitude could be calculated as

$$\|\Delta g(x, y)\| = \sqrt{(g_x(x, y))^2 + (g_y(x, y))^2}. \quad (4.2)$$

The gradient orientation is unsigned, it is defined as

$$\theta(x, y) = \arctan \frac{g_y(x, y)}{g_x(x, y)} + \frac{\pi}{2}. \quad (4.3)$$

To compute the gradient histogram of a cell, each pixel casts a vote weighted by its gradient magnitude, for the bin corresponding to its gradient orientation. All of gradient orientations are group into 9 bins. Thus, every block has a gradient histogram with 36 dimensions, and ROI-HOG feature vector has $49 * 36 = 1764$ dimensions. Furthermore, integral histograms of oriented gradients (IHOG) [23] are utilized for farther speed up the process of feature extraction. The histograms of oriented gradients of the pixel (x, y) could be expressed as follows:

$$T(x, y) = [g_1, \dots, g_i, \dots, g_9], \quad g_i = \begin{cases} \|\Delta g(x, y)\|, & i = k \\ 0, & i \neq k, \end{cases} \quad (1 \leq i \leq 9; 1 \leq k \leq 9). \quad (4.4)$$

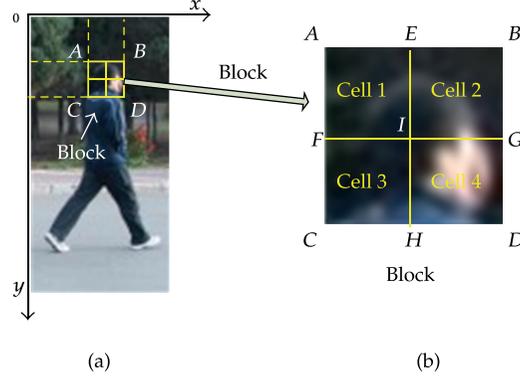


Figure 7: IHOG calculation of a block.

The integral feature vectors in x -orientation is as follows:

$$H(x, y) = \sum_{i=0}^x T(i, y). \quad (4.5)$$

The integral feature vectors in y -orientation is as follows:

$$HH(x, y) = \sum_{i=0}^y H(x, i). \quad (4.6)$$

As shown in Figure 7, IHOG of a cell could be calculated as

$$\text{HOG}_{\text{CELL-1}} = HH(A) + HH(I) - HH(E) - HH(F). \quad (4.7)$$

Accordingly, IHOG of a block could be calculated as

$$\text{HOG}_{\text{BLOCK}} = [\text{HOG}_{\text{CELL-1}}, \text{HOG}_{\text{CELL-2}}, \text{HOG}_{\text{CELL-3}}, \text{HOG}_{\text{CELL-4}}]. \quad (4.8)$$

IHOG method only need scan the entire image for once and storage the integral gradient data. Any area's HOG feature could be obtained with simple addition and subtraction operations without repeated calculation of the gradient orientation and magnitude of each pixel.

4.2. Sample Selection for Training

For pedestrian recognition in urban road environment, we build a pedestrian sample dataset called RSPerson (Person Dataset of Road System) dataset. In the sample dataset, the positive samples are including walking pedestrians, standing still pedestrians, and group pedestrians with different size, pose, gait, and clothing. Some preexperimental studies have shown that the selection of negative samples is particularly important for reduction of false alarms. Thus,

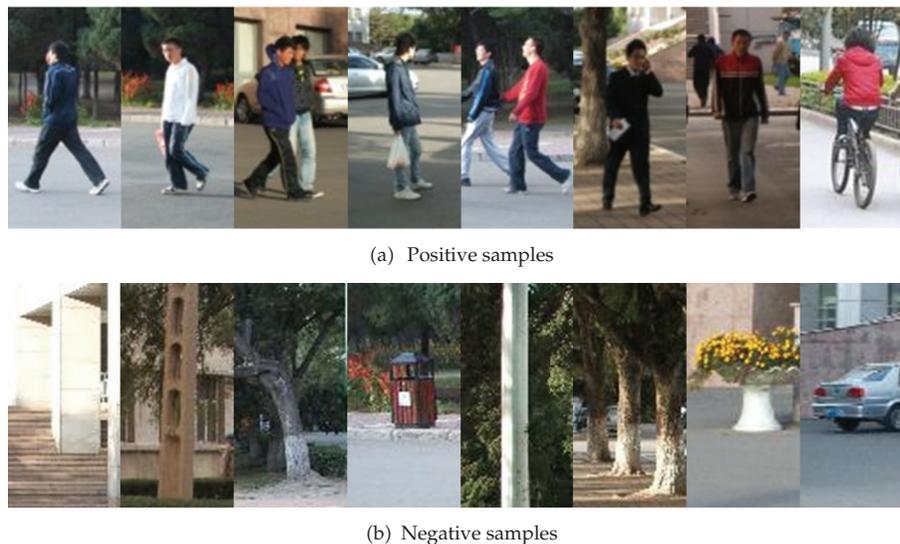


Figure 8: Partial samples in RSPerson dataset.

boles, trash cans, telegraph poles, and bushes which are likely to be mistaken for pedestrians, as well as some normal objects such as roads, vehicles, and other infrastructures are selected to form negative samples. This is most beneficial for our pedestrian detection system. In RSPerson dataset, each sample image is normalized to 64×128 pixels for training. Figure 8 shows some samples of RSPerson dataset.

4.3. Pedestrian Recognition with SVM

Before online recognizing pedestrian, we should construct a classifier offline trained by SVM algorithms. Firstly, training dataset and test dataset are built from RSPerson dataset. The training dataset includes 2000 pedestrian and 2000 nonpedestrian samples, and the testing dataset includes 500 pedestrian and 500 nonpedestrian samples. The training dataset samples are handled, and features are extracted to form training vectors. With cross-validation based on grid search method, the proper parameters of SVM are selected. RBF kernel is chosen as kernel function, and the penalty factor $C = 1024$ as well as the kernel parameter $g = 0.0625$. After that, the pedestrian classifier could be constructed. Finally, testing dataset samples are chosen to test the performance of the classifier. We use the DET curve which contains two indicators: miss rate and FPPW (false positive per window) to evaluate performance of SVM classifiers. The performance of pedestrian recognition based on ROI-IHOG is shown in Figure 9.

For online recognition, once the potential pedestrian locations are located by laser radar, candidate regions in the image are confirmed accordingly by the perspective mapping model. For each candidate region, scale transforming is carried out for normalization of $64 * 128$ pixels, and then, ROI-IHOG feature vector could be extracted. Based on these steps, we can judge whether the candidate is a true pedestrian or not by the classifier trained with SVM.

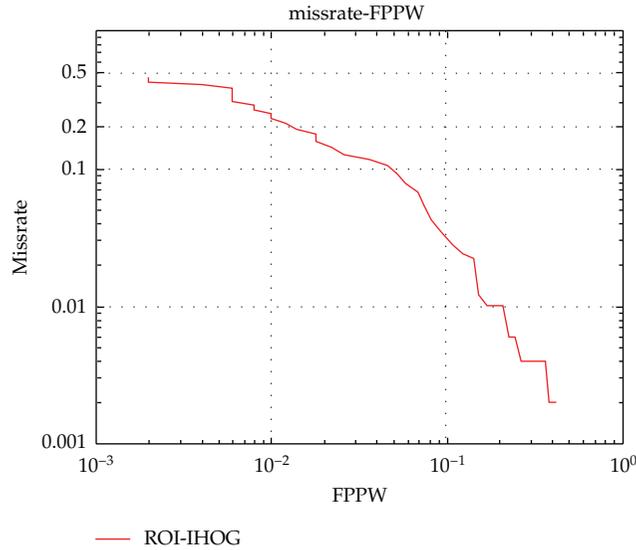


Figure 9: Performance on ROI-IHOG-based pedestrian recognition.

5. Experimental Results

For testing the validity of the proposed method, several video sequences from realistic urban traffic scenarios are tested for performance assessment of our pedestrian recognition experimental platform. Firstly, the pedestrian candidate locations are estimated based on laser radar data processing and space-image perspective mapping model. Some candidate region segmentation results are shown in Figure 10. In this way, potential pedestrian regions are located in the image, but some other obstacles (poles, shrub, etc.) are also located as positives.

Secondly, the proposed ROI-IHOG+SVM algorithm is tested with several video sequences. In this step, pedestrian recognition only depends on ROI-IHOG+SVM for an entire image without fusing the laser information. The recall could reach 93.8% under 10^{-4} FPPW. The image size is $320 * 240$ pixels. The average detection time is about 600 ms/frame. Some detection results are shown in Figure 11.

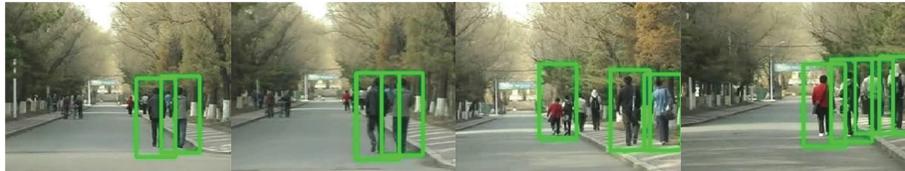
Finally, fusing information from laser and vision sensor, each candidate region detected is scaled to the size of $64 * 128$ pixels and extracted the ROI-IHOG feature. According to our recognition method, the candidate region is considered to be a pedestrian or not by the classifier trained with SVM. Based on multisensor fusion, the average detection time is about 18 ms for a candidate. Thus, if there are 5 candidate regions in each image of the video sequence averagely, the processing speed is about 11 frame/s which could be satisfied the real time requirement. Several recognition results (Figure 12) indicate that the proposed pedestrian detection approach based on multisensor fusion has good performance, which could provide an effective support for active pedestrian safety protection.

6. Conclusions

A fast pedestrian recognition algorithm based on multisensor fusion is developed in this paper. Potential pedestrian candidate regions are located by laser scanning and the perspective mapping model, and then ROI-IHOG feature extraction method is proposed



Figure 10: Pedestrian candidate region estimation results under different urban scenarios.

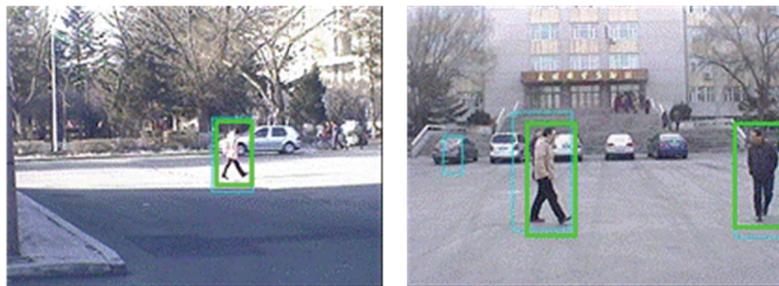


(a) 1st urban scenario



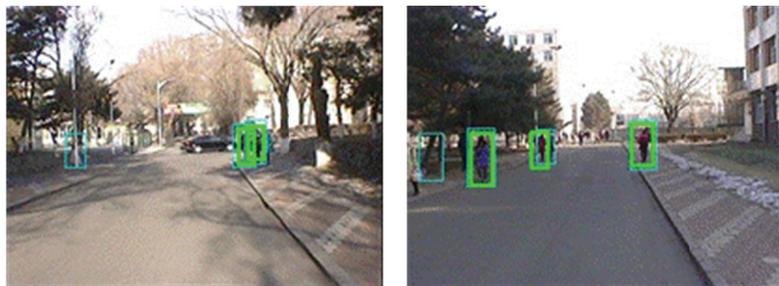
(b) 2nd urban scenario

Figure 11: Pedestrian detection with ROI-IHOG+SVM.



(a)

(b)



(c)

(d)

Figure 12: Pedestrian recognition based on multisensor fusion.

for reducing computational time cost. Moreover, SVM is utilized with a novel pedestrian sample dataset which adapt to the urban road environment for online recognition. Pedestrian recognition is tested with radar, vision, and two-sensor fused, respectively. Reliable and timewise performances are shown on fusion-based pedestrian recognition. The processing speed could reach 11 frame/s which could be satisfied the real time requirement. In future work, we will further study the key technologies for pedestrian safety, such as pedestrian tracking, pedestrian behavior recognition, and conflict analysis between pedestrians and the host vehicle.

Acknowledgments

This work is partly supported by National Science Foundation of China (nos. 51108208, 51278220), Postdoctoral Science Foundation Funded Project of China (no. 20110491307), and Fundamental Research Funds for the Central Universities of China (no. 201103146).

References

- [1] L. Guo, P.-S. Ge, M.-H. Zhang, L.-H. Li, and Y.-B. Zhao, "Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4274–4286, 2012.
- [2] W. Wang, W. Zhang, H. Guo et al., "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research Part C-Emerging Technologies*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [3] T. Gandhi and M. M. Trivedi, "Pedestrian protection systems: issues, survey, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 3, pp. 413–430, 2007.
- [4] D. M. Gavrilu and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *International Journal of Computer Vision*, vol. 73, no. 1, pp. 41–59, 2007.
- [5] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *International Journal of Computer Vision*, vol. 63, no. 2, pp. 153–161, 2005.
- [6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, San Diego, Calif, USA, 2005.
- [7] O. Tsimhoni, M. J. Flannagan, and T. Minoda, "Pedestrian detection with night vision systems enhanced by automatic warnings," Tech. Rep. UMTRI-2005-23, 2005.
- [8] E. Binelli, A. Broggi, A. Fascioli et al., "A modular tracking system for far infrared pedestrian recognition," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '05)*, pp. 759–764, June 2005.
- [9] K. C. Fuerstenberg and J. Scholz, "Reliable pedestrian protection using laserscanners," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '05)*, pp. 142–146, Las Vegas, Nev, USA, June 2005.
- [10] A. Ewald and V. Willhoeft, "Laser scanners for obstacle detection in automotive applications," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '00)*, pp. 682–687, Dearbon, Mish, USA, 2000.
- [11] E. Moxey, N. Johnson, and M. G. McCarthy, "The advanced protection of vulnerable road users (APVRU)—project information sheet," *Target Tracking: Algorithms and Applications*, 2004.
- [12] Y. L. Guilloux, J. Lonnoy, and R. Moreira, "PAROTO Project: the benefit of infrared imagery for obstacle avoidance," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '02)*, June 2002.
- [13] D. M. Gavrilu, M. Kunert, and U. Lages, "A multi-sensor approach for the protection of vulnerable traffic participants—the PROTECTOR project," in *Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference*, pp. 2044–2048, May 2001.
- [14] D. M. Gavrilu, J. Giebel, and S. Munder, "Vision-based pedestrian detection: the PROTECTOR system," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '04)*, pp. 410–415, June 2004.
- [15] M. M. Michael and O. M. Andrzej, "Enhancing pedestrian safety by using the SAVE-U pre-crash system," in *Proceedings of the IEEE 12th World Congress on ITS*, San Francisco, Calif, USA, 2005.
- [16] P. Marcha, "SAVE-U: sensors and system Architecture for Vulnerable road Users protection," in *Proceedings of the 3rd Concertation meeting*, Bruxelles, Belgium, 2004.

- [17] U. Scheunert, H. Cramer, B. Fardi, and G. Wanielik, "Multi sensor based tracking of pedestrians: a survey of suitable movement models," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '04)*, pp. 774–778, June 2004.
- [18] M. Szarvas, U. Sakait, and J. Ogata, "Real-time pedestrian detection using LIDAR and convolutional neural networks," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '06)*, pp. 213–218, June 2006.
- [19] M. Töns, R. Doerfler, M. M. Meinecke, and M. A. Obojski, "Radar sensors and sensor platform used for pedestrian protection in the EC-funded project SAVE-U," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '04)*, pp. 813–818, June 2004.
- [20] M. Bertozzi, A. Broggi, M. Felisa, G. Vezzoni, and M. Del Rose, "Low-level pedestrian detection by means of visible and far infra-red tetra-vision," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '06)*, pp. 231–236, 2006.
- [21] A. Broggi, P. Cerri, S. Ghidoni, P. Grisleri, and H. G. Jung, "A new approach to urban pedestrian detection for automatic braking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 4, pp. 594–605, 2009.
- [22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 886–893, June 2005.
- [23] F. Porikli, "Integral histogram: a fast way to extract histograms in Cartesian spaces," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 829–837, June 2005.

Research Article

A Novel Short-Range Prediction Model for Railway Track Irregularity

Peng Xu, Rengkui Liu, Quanxin Sun, and Futian Wang

*MOE Key Laboratory for Urban Transportation Complex Systems Theory, Beijing Jiaotong University,
No. 3 Shang Yuan Cun, Haidian District, Beijing 100044, China*

Correspondence should be addressed to Peng Xu, suepen.aili@126.com

Received 27 August 2012; Revised 12 November 2012; Accepted 15 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Peng Xu 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 recent years, with axle loads, train loads, transport volume, and travel speed constantly increasing and railway network steadily lengthening, shortcomings of current maintenance strategies are getting to be noticed from an economical and safety perspective. To overcome the shortcomings, permanent-of-way departments throughout the world have given a considerable attention to an ideal maintenance strategy which is to carry out appropriate maintenances just in time on track locations really requiring maintenance. This strategy is simplified as the condition-based maintenance (CBM) which has attracted attentions of engineers of many industries in the recent 70 years. To implement CBM for track irregularity, there are many issues which need to be addressed. One of them focuses on predicting track irregularity of each day in a future short period. In this paper, based on track irregularity evolution characteristics, a Short-Range Prediction Model was developed to this aim and is abbreviated to TI-SRPM. Performance analysis results for TI-SRPM illustrate that track irregularity amplitude predictions on sampling points by TI-SRPM are very close to their measurements by Track Geometry Car.

1. Introduction

Traffic accidents have long been a social-economical problem which has caused increasing concerns to the public worldwide [1, 2]. According to statistics on train accidents by Office of Safety of US Federal Railroad Administration, 542933 people were injured or killed by railway accidents mainly resulting from railway track from January 1975 to May 2011 [3].

The roles of railway track are to guide the vehicle or locomotive traveling, endure comprehensive effects of contributing environmental variables and loads from wheels, distribute wheel loads to roadbed, bridges, or tunnels, provide a smooth trajectory for wheel running, keep operation costs, travel safety and passenger comfort within an accepted range, and roadbed in a good condition [4, 5]. Track condition may be classified into two categories:

track geometry condition and track structure condition [6]. Track geometry condition plays a significant role in ensuring train traveling safety and passenger comfort [7] and is usually measured by several geometry parameters [5]. These parameters comprise projections of rails in the horizontal, longitudinal and vertical planes and include gauge and cross level in the horizontal plan, longitudinal level in the vertical plan, alignment in the longitudinal plane, and twist for measuring uneven running surface [8]. In theory, values of these five parameters that is, gauge, cross level, longitudinal level, alignment, and twist are G_0 which is the designed gauge and in most countries is equal to 1435 mm, 0 mm, 0 mm, 0 mm, and 0 mm/m for the straight track, respectively, and for the curved track are $G_0 + G_w$, wherein G_w is the designed gauge widening, and CL_0 which is the designed super-elevation, 0 mm, 0 mm and 0 mm/m, respectively. Because of many reasons [4, 9, 10], however, the actual values of these parameters deviate from their theoretical ones. These deviations are usually referred to as track irregularity [5]. The existence of track Irregularity makes unsmooth the running trajectory of wheels. The unsmooth trajectory will lead to vibrations of vehicles and locomotives and dynamic forces between wheels and rails. The trains vibrations and dynamic forces not only reduce train travelling safety and passenger comfort, but also shorten track equipment life [4, 11], what is more, deteriorating track irregularity further [4, 12]. According to investigation reports, which were published by the Rail Accident Investigation Branch of UK, of 184 accidents occurred between October 2005 and June 2010, there are 58 derailments, and 24 of them are induced by bad track irregularity [13].

In recent years, with axle loads, train loads, transport volume, and travel speed constantly increasing and railway network steadily lengthening, shortcomings of current maintenance strategies are getting to be noticed from an economical and safety perspective [9]. To overcome the shortcomings, permanent-of-way departments throughout the world have given a considerable attention to an ideal maintenance strategy which is to carry out appropriate maintenances just in time on track locations requiring maintenance. This strategy is simplified as the condition-based maintenance (CBM) which has attracted attentions of engineers of many industries in the recent 70 years. To implement CBM strategy for track irregularity, there are many issues which need to be addressed [9]. One of them deals with predicting track irregularity of each day in the future short period.

There are some studies on track irregularity prediction methods. Professionals at Railway Technical Research Institute in Japan have, respectively, developed degradation models which utilize standard deviations of track geometry measurements to predict standard deviations of track alignment and surface over 100-meter-long track sections [14]. Alfelor et al. have built a track degradation database storing track gauge restraints and track geometry parameters measured with Gage Restraint Measurement System, traffic loads, environment, and track structural characteristics, and they developed a track degradation analysis program which can formulate a one-to-one relationship between track degradation and a contributing parameter by employing least square linear regression [15]. Chen et al. at China Academy of Railway Sciences have developed an integrated factor method (IFM) which utilizes the latest track geometry measurements to predict track irregularity parameter values on all sampling points [10]. In order to make predictions, IFM divides a railway track into 200-meter-long unit sections, classifies unit sections into 17 categories, and assumes that unit sections in the same category have the same track irregularity deterioration rate. Based on a common denominator; "a good track behaves well, while a poorer one deteriorates faster," Veit and Marschnig developed an exponential model for describing track irregularity evolution between two adjacent temping maintenances over a 5-meter-long track section [12]. Quiroga and Schnieder also developed an exponential model for describing track surface

evolution between two adjacent temping maintenances over 200-meter-long sections [7]. Based on experiences of railway engineers, Meier-Hirmer et al. employed Gamma process to fit track surface deterioration rate between two adjacent temping maintenances over track sections with the length equal to 200 meters and 1000 meters [16]. Xu et al. developed a multistage framework integrating linear regression models to describe track irregularity evolution between two adjacent maintenances [17]. And based on conclusions drawn from [9, 16], Xu et al. then proposed a novel method for predicting track irregularity condition over unit sections [18] and mean of track irregularity amplitudes over 10-meter-long unit sections [19].

There are some limitations of these methods [9]. The study presented in this paper will overcome these shortcomings and develop a novel, distinctive model to predict track irregularity parameter values over sampling points of each day within the following one or two inspection intervals of Track Geometry Car (TGC). Thus, the model is named Short-Range Prediction Model for Track Irregularity and is abbreviated to TI-SRPM.

The rest of the paper is organized as follows. Section 2 declares variables for TI-SRPM. In Section 3, TI-SRPM is detailed. In Section 4, by analyzing track gauge and surface predictions of about 8000 sampling points, the performance of TI-SRPM is analyzed. Based on the performance analysis results, conclusions are drawn in Section 5.

2. Variable Denotations

Variable denotations throughout this paper are listed as follows.

Sec: a track section on which TI-SRPM is built.

Ti: a track irregularity index for which TI-SRPM is built.

$\{t_{n,x}\}$: a date series of every day within two adjacent maintenances at Sec; when x is equal to 0, $t_{n,0}$ presents the date when a TGC inspection was carried out; when x is more than 0, $t_{n,x}$ presents the date of the x th day after the n th TGC inspection; when x is less than 0, $t_{n,x}$ presents the date of the x th day before the n th TGC inspection.

$a_{n,x}^l$: the measurement by TGC of the amplitude of the track irregularity index Ti on $t_{n,x}$ over the sampling point l in Sec; due to the fact that TGC is not operated to measure track irregularity each day, only $a_{n,0}^l$ has the actual value.

$\hat{a}_{n,x}^l$: the prediction of $a_{n,x}^l$.

$$A_n = \begin{pmatrix} a_{n,0}^1 & \cdots & a_{n,x}^1 \\ \vdots & \ddots & \vdots \\ a_{n,0}^l & \cdots & a_{n,x}^l \end{pmatrix},$$

$$\hat{A}_n = \begin{pmatrix} \hat{a}_{n,1}^1 & \cdots & \hat{a}_{n,x}^1 \\ \vdots & \ddots & \vdots \\ \hat{a}_{n,1}^l & \cdots & \hat{a}_{n,x}^l \end{pmatrix},$$

$$\begin{aligned}
A_{n,x} &= [a_{n,x}^1, \dots, a_{n,x}^l]^T, \\
\widehat{A}_{n,x} &= [\widehat{a}_{n,x}^1, \dots, \widehat{a}_{n,x}^l]^T, \\
A_n^l &= [a_{n,1}^l \dots a_{n,x}^l], \\
\widehat{A}_n^l &= [\widehat{a}_{n,1}^l \dots \widehat{a}_{n,x}^l].
\end{aligned} \tag{2.1}$$

S_n : the condition of Ti over Sec on $t_{n,0}$.

k_n : the degradation rate of Ti within a date interval including the date of the n th TGC inspection.

$V_l = \underbrace{[1, 1, \dots, 1]}_l^T$, a column vector with the size of l and all elements of 1.

3. TI-SRPM

Many factors exert influences on track irregularity deteriorations. To ensure TI-SRPM to capture track irregularity degradation trend as much as possible, it should be built on the basis of track irregularity degradation characteristics.

3.1. TI-SRPM Research Object

According to practical experiences on track irregularity degradation, under the effect of many factors on track irregularity deterioration, degradation processes of two near spatial points are different, and what is more, degradation processes of two track irregularity indices over a spatial point are also distinguished. On the other hand, maintenance works routinely or temporarily scheduled will destroy the current degradation process and initiate a new evolution process. Therefore, track irregularity prediction model should be built for the degradation process of a track irregularity index over a spatial point within two adjacent maintenance works.

However, because there are mileage errors existing between TGC inspections [9, 20, 21], building track irregularity prediction model on a spatial point is usually impossible. Therefore, some researches on track irregularity predictions built models for a section of some length.

Track Geometry Measurements (TGMs) used in this paper are processed with two mileage error correction models, Key Equipment-based Mileage Error Correction model, KE-BMEC [9, 22], and Track Geometry Measurements-based Mileage Error Correction model, TGM-BMEC [9]. Mileage errors of such TGM processed by these two models are normally less than a sampling distance, that is, 0.25 m. Therefore, TI-SRPM will be constructed for the degradation process of a track irregularity index Ti over a shorter unit section, 0.5 m. Due to the fact that a unit section only has 2~3 sampling points, amplitudes of Ti over these points do not deviate very much from each other. Therefore, track irregularity condition, S_n , of Ti over Sec on $t_{n,0}$ is quantified with the mean of track irregularity measurements of Ti over sampling points in Sec.

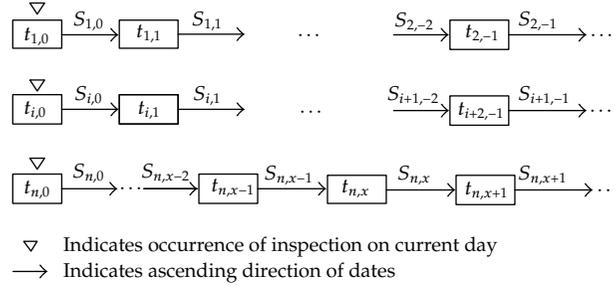


Figure 1: Evolutionary process of track irregularity over the unit section Sec.

3.2. Track Irregularity Degradation Process of T_i over the Unit Section Sec

With the current track irregularity condition as its initial condition, future track irregularity condition is induced by comprehensive influences of many factors on track irregularity. During railway transportation, track irregularity changes on a daily basis. The difference in track irregularity condition on two moments quantifies accumulatively comprehensive effects in the period. It indicates that comprehensive influences of factors on track irregularity deteriorations can be taken into consideration by directly using TGM. This forms a basis for some pieces of research on track irregularity predictions, for example, IFM in [9].

Therefore, $s_{n,x-1}$ on $t_{n,x-1}$ is a main input affecting $s_{n,x}$ on $t_{n,x}$. Meanwhile, $s_{n,x}$ serves as a main output at $t_{n,x}$ which influences $s_{n,x+1}$ on $t_{n,x+1}$. Figure 1 illustrates this evolution process. In the evolution process of Figure 1, track irregularity condition is inspected by TGM at an inspection cycle. Track irregularity condition $s_{n,x}$ on the day of t_n can be calculated from TGM generated during this day's inspection.

3.3. TI-SRPM Building

Research results of [8] show that in a normal circumstance (no natural disasters occurring) and with no maintenance works carried out, track irregularity condition changes over a unit section within a short period are small. On the other hand, under effects of factors, changes in track irregularity condition over a unit section are nonlinear with railway transportation [7, 9–11, 14, 16–19]. So, in the normal circumstance, the track irregularity degradation process can be presented as an approximately differentiable and evolutionarily smooth curve. According to one of calculus fundamentals, a nonlinear deterioration process in a short-range period within two adjacent maintenance works can be approximated by a linear deterioration process in the same short period. A whole evolution process between two adjacent maintenance works comprises deterioration processes of many short periods. Therefore, the whole process can be approximated by many linear deterioration processes.

According to the previous discussion, once the short period lengths are known for the approximation of a nonlinear deterioration process over Sec by many linear processes, TGM can then be used to fit these linear deterioration processes, and a mean deterioration rate, k_q , is available for each of these linear processes. k_q quantifies the mean comprehensive effect of many factors on the condition of T_i over Sec. Because the unit section has a shorter length, all sampling points in Sec can be considered to have the same deterioration rate, k_q , for the track

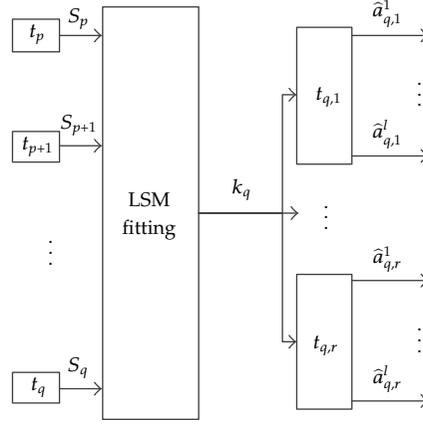


Figure 2: Illustration of predicting track irregularity amplitudes over all sampling points in the unit section Sec after the q th Track Geometry Car inspection.

irregularity index Ti. According to the analysis results of Section 3.2, the amplitude of Ti over a sampling point l on $t_{q,x}$ is approximated by $a_{q,0}^l + k_q \cdot x$.

After the q th TGC inspection for the railway track where the unit section Sec is, predicting amplitudes of Ti at each day within the period of $t_{q,1} \sim t_{q,r}$ over sampling points in Sec is illustrated in Figure 2. In Figure 2, the date range satisfies the following two conditions: (a) $t_{q,r} - t_p = T_q$, wherein T_q represents the short period length for the approximation of the current nonlinear deterioration process, and (b) in the period of $t_{q,1} \sim t_{q,r}$, there is at least one TGC inspection for the railway track where Sec is. r is determined by the TGC inspection interval and usually takes on the maximum interval, I_d . In Figure 2, the prediction for amplitudes $A_{q,x}$ of Ti on $t_{q,x}$ over sampling points in Sec is made with (3.1), wherein the mean deterioration rate, k_q , is determined by fitting the time series $\{S_p, S_{p+1}, \dots, S_q\}$ with the least squares method, where

$$\hat{A}_{q,x} = A_{q,0} + k_q \cdot x \cdot V_l. \quad (3.1)$$

Because the track irregularity deterioration is nonlinear, to approximate the real deterioration process with a linear process, when a new TGC inspection for the railway track is carried out, the newly generated TGM should be taken into consideration to adjust previously fitted deterioration rate, k_q , and to obtain a new one, k_{q+1} . The new deterioration rate is then used for making predictions as shown in Figure 3. In Figure 3, the date range satisfies the same conditions with one in Figure 2. Predictions for amplitudes, $A_{q+1,x}$ of Ti on $t_{q,x+1}$ in Sec are made according to (3.2). The mean deterioration rate in (3.2) is obtained according to the newly formed time series $\{S_u, S_{u+1}, \dots, S_{q+1}\}$ where

$$\hat{A}_{q+1,x} = A_{q+1,0} + k_{q+1} \cdot x \cdot V_l \quad (3.2)$$

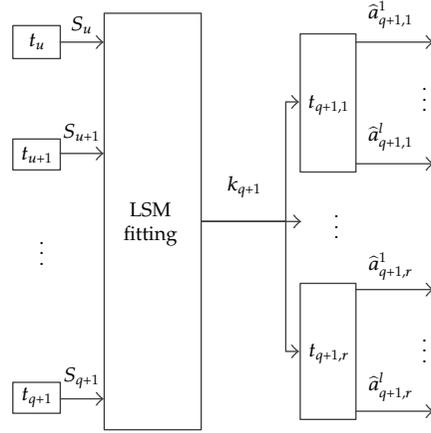


Figure 3: Illustration for predicting track irregularities of all sampling points along the unit section Sec after the $(q + 1)$ th Track Geometry Car inspection.

As the TGC inspection for the railway track where Sec is continues, the previous adjustments will list a family of linear prediction models, shown in (3.3). The family of linear models is TI-SRPM for making short-range predictions for amplitudes of Ti over sampling points in Sec consider

$$\begin{aligned}
 \hat{A}_{2,x} &= A_{2,0} + k_2 \cdot x \cdot V_l \\
 &\vdots \\
 \hat{A}_{q,x} &= A_{q,0} + k_q \cdot x \cdot V_l \\
 &\vdots
 \end{aligned}
 \tag{3.3}$$

In (3.3), x lies in the closed section $[1, r]$, and the mean deterioration rate, k_q , for each linear model is determined with the least squares methods by fitting $\{S_p, S_{p+1}, \dots, S_q\}$ calculated from TGM generated during the least 4 TGC inspections. When the available TGC inspections are less than 4 in times, all available TGM are used.

4. Performance Analysis

Utilizing TGM processed with both KE-BMEC and TGM-BMEC, multiple predictions for amplitudes of track irregularity over sampling points on the Jiulong-Beijing Railway Track under the jurisdiction of Jinan Railway Bureau have been made with TI-SRPM. In this section, errors in predictions of amplitudes of gauge and right surface over sampling points in the section from K612+000 to K614+000 will be analyzed. There are about 4000 $(2000/0.5)$ unit sections and about 8000 $(2000/0.25)$ sampling points in the 2-kilometer-long track section. To perform these predictions, approximately 48000 $(4000*6*2)$ linear models in the form of (3.1) or (3.2) are built automatically with a computer routine coded for TI-SRPM.

Table 1: Some error statistics from gauge predictions of all sampling points in Jiulong-Beijing track section from K612+000 to K614+000.

Date	Error statistic		
	ME (mm)	SD (mm)	ρ
Sep. 24	-0.0018	0.1912	0.9884
Oct. 10	-0.0019	0.0660	0.9985
Oct. 30	-0.0013	0.3007	0.9690
Nov. 13	-0.0022	0.1062	0.9960
Dec. 12	-0.0027	0.1481	0.9921
Dec. 25	0.0001	0.1030	0.9961

These predictions of 6 times were made after inspections of September 8, September 24, October 10, October 30, November 13, and December 12, 2008, respectively. As stated in Section 3, due to the fact that TGC does not perform inspections every day, track irregularity measurements are only available for those days with inspection. Thus, error analysis is carried out only for these predictions which have amplitude measurements, that is, on September 24, October 10, October 30, November 13, December 12, and December 25, 2008.

4.1. Gauge

Histograms of errors in gauge predictions are illustrated in Figure 4. And some statistics from gauge prediction errors such as Mean Error, ME, Standard Deviation, SD, and Correlation Coefficient, ρ , of errors are listed in Table 1.

Histograms in Figure 4 and statistics listed in Table 1 show that gauge prediction errors of these 6 times are normally distributed with means approximately equal to 0 mm, the maximum SD is 0.3007 mm, and the minimum ρ takes the value of 0.9690.

From the previous results, the following four main conclusions are arrived at: (a) the gauge deterioration trends over sampling points are captured by TI-SRPM, (b) gauge amplitude measurements, $a_{n,x}^l$, lie in the closed section $[\hat{a}_{n,x}^l - 2SD, \hat{a}_{n,x}^l + 2SD]$ (wherein, $SD < 0.31$ mm) with the probability of 95.44%, (c) gauge predictions are very close to gauge measurements, and (d) there are high similarities between gauge measurements and predictions.

4.2. Right Surface

Figure 5 shows histograms of right surface prediction errors. Some error statistics are listed in Table 2.

Histograms in Figure 5 and error statistics listed in Table 2 show that right surface prediction errors are normally distributed with means approximately equal to 0 mm, the maximum SD is 0.1487 mm, and the minimum ρ takes 0.9940.

From the previous results, the similar conclusions with ones in Section 4.1 are drawn as follows: (a) right surface deterioration trends are captured by TI-SRPM, (b) right surface measurements, $a_{n,x}^l$, lie in the closed section $[\hat{a}_{n,x}^l - 2SD, \hat{a}_{n,x}^l + 2SD]$ (wherein, $SD < 0.15$ mm) with the probability of 95.44%, (c) right surface predictions are very close to right surface measurements, and (d) there are high similarities between right surface measurements and predictions.

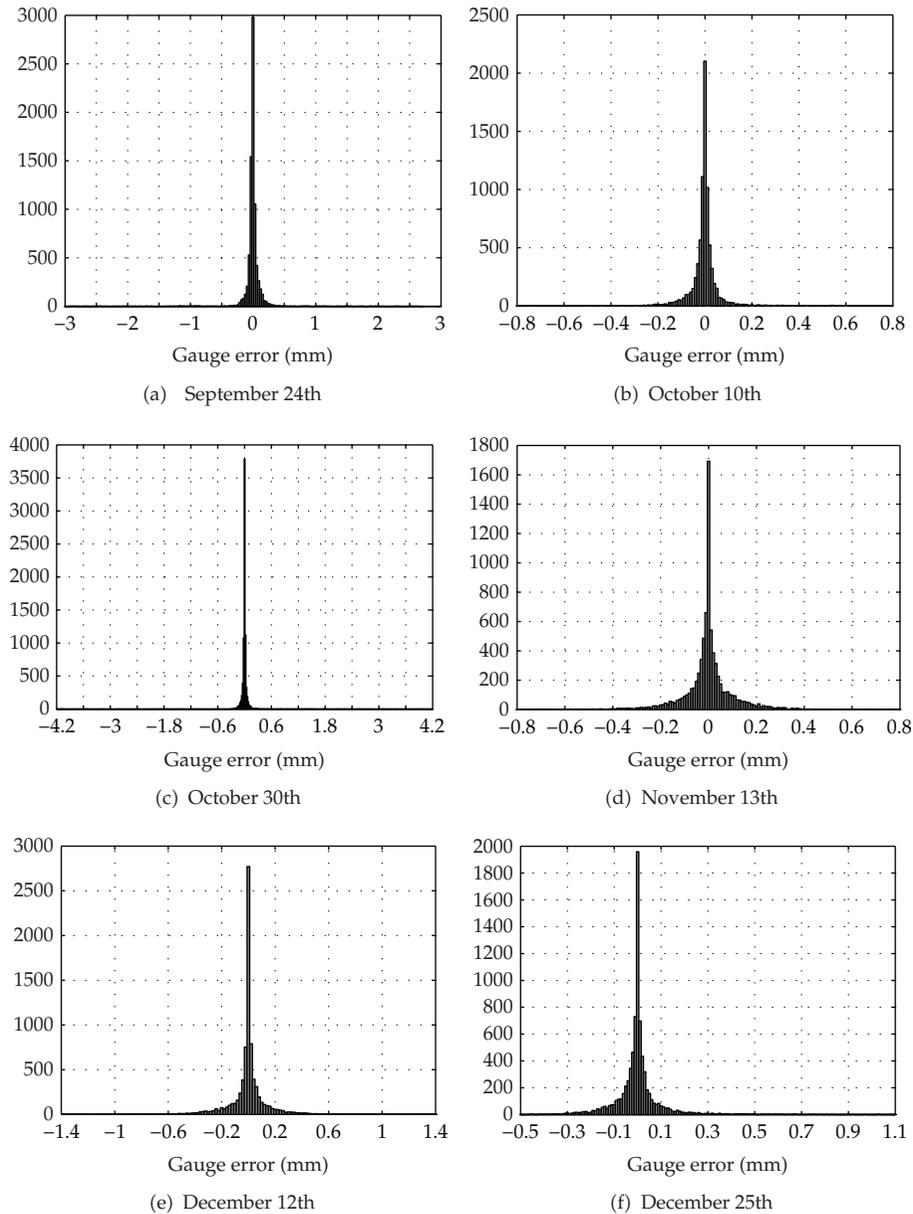


Figure 4: Histograms on errors in gauge predictions of all sampling points in Jiulong-Beijing track section from K612+000 to K614+000.

5. Conclusions

According to one of calculus fundamentals, on the basis of track irregularity deterioration characteristics, a short-range prediction model was developed for amplitudes of track irregularity over sampling points on each day within a future short period. The research objects of the model are a deterioration process of a track irregularity parameter between two adjacent maintenance works over a 0.5-meter-long unit section. TI-SRPM for a research

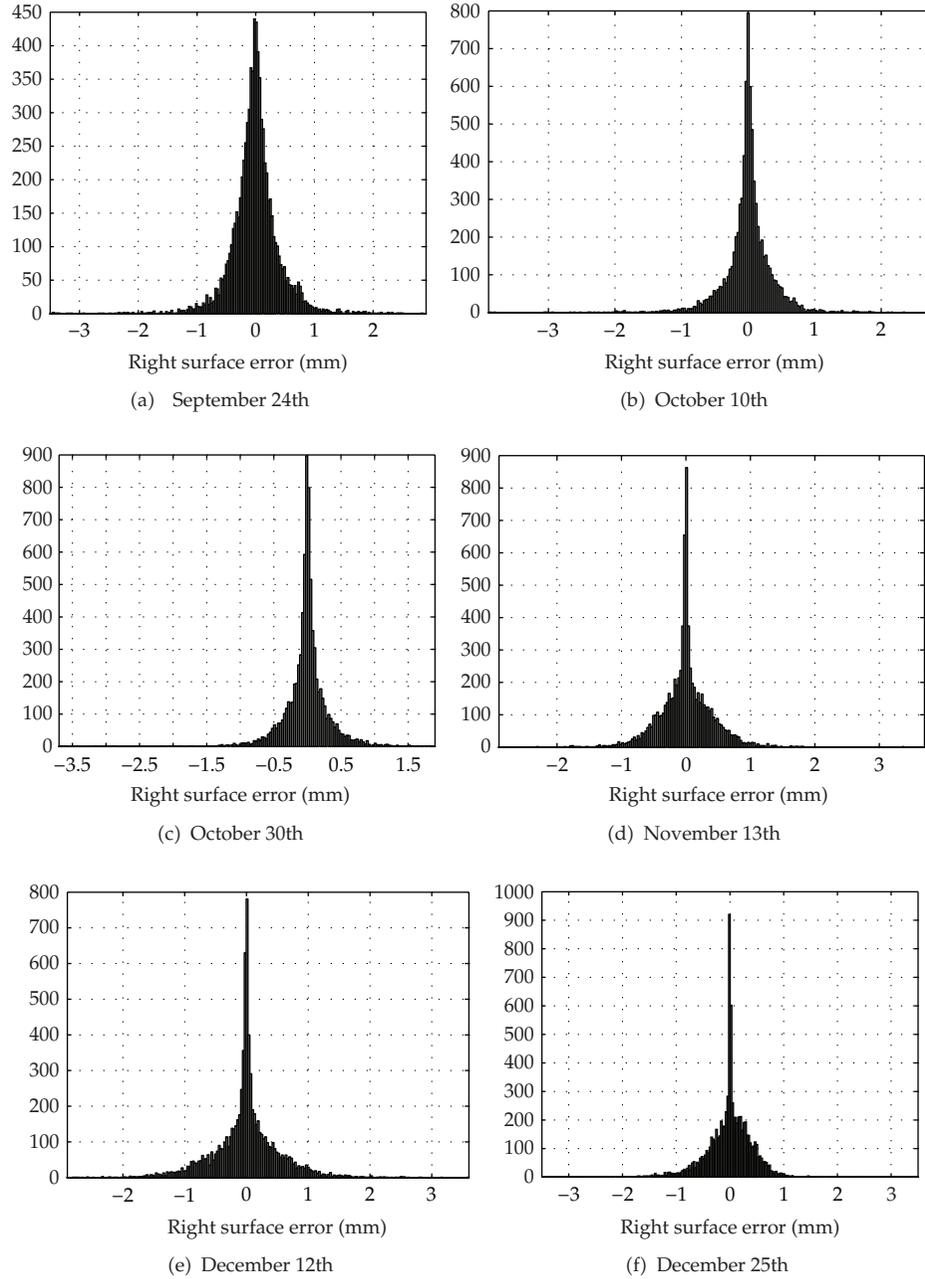


Figure 5: Histograms on errors in right surface predictions of all sampling points in Jiulong-Beijing track section from K612+000 to K614+000.

object comprises a family of linear models, that is, $\hat{A}_{q,x} = A_{q,0} + k_q \cdot x \cdot V_l$ (wherein, $q \geq 2$ and $x \in [1, r]$). Performance analysis in Section 4 for TI-SRPM shows that track irregularity deterioration trends over sampling points are captured by TI-SRPM, and track irregularity predictions are very close to their measurements.

Table 2: Some error statistics from right surface predictions of all sampling points in Jiulong-Beijing track section from K612+000 to K614+000.

Date	Error statistic		
	ME (mm)	SD (mm)	ρ
Sep. 24	-0.0047	0.1477	0.9940
Oct. 10	-0.0015	0.0931	0.9977
Oct. 30	-0.0010	0.0976	0.9975
Nov. 13	0.0019	0.1378	0.9949
Dec. 12	-0.0034	0.1486	0.9943
Dec. 25	0.0005	0.0735	0.9986

Acknowledgments

This research was partly sponsored by the National Key Technology R&D Program (2009BAG12A10) and the National Basic Research Program of China (2012CB725406).

References

- [1] H. W. Guo, W. H. Wang, W. W. Guo, X. B. Jiang, and B. Heiner, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [2] W. H. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, Communications Press, Beijing, China, 2012.
- [3] Federal Railroad Administration Office of Safety Analysis-Train Accidents, <http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/inctally1.aspx>.
- [4] S. Yoshihiko, *New Railway Track Dynamics*, China Railway Press, Beijing, China, 2001.
- [5] C. Esveld, *Modern Railway Track*, Delft University of Technology Publishing Services, Delft, The Netherlands, 2nd edition, 2001.
- [6] J. M. Sadeghi and H. Askarinejad, "Development of track condition assessment model based on visual inspection," *Structure and Infrastructure Engineering*, vol. 7, no. 12, pp. 895–905, 2011.
- [7] L. M. Quiroga and E. Schnieder, "A heuristic approach to railway track maintenance scheduling," in *Proceedings of the 12th International Conference on Computer System Design and Operation in the Railways and other Transit Systems (COMPRAIL '10)*, pp. 687–699, WIT Press, Beijing, China, September 2010.
- [8] CEN, *prEN13848-1:2003 Railway Applications-Track-Track Geometry Quality-Part 1: Characterization of Track Geometry*, CEN, Brussels, Belgium, 2004.
- [9] P. Xu, *Mileage error correction model for track geometry data from track geometry car & track irregularity prediction model [dissertation]*, Beijing Jiaotong University, Beijing, China, 2012.
- [10] X. M. Chen, L. Wang, F. C. Yang, X. S. Chai, and W. Q. Wu, "Integrating factor method for predicting the developing trend of railway track irregularity," *China Railway Science*, vol. 27, no. 6, pp. 27–31, 2006 (Chinese).
- [11] J. J. Fan, *Modern Railway Track*, China Railway Press, Beijing, China, 2nd edition, 2004.
- [12] P. Veit and S. Marschnig, "Sustainability in track—a precondition for high speed traffic," in *Proceedings of ASME Joint Rail Conference (JRC' 10)*, pp. 349–355, Urbana, Ill, USA, April 2010.
- [13] Rail Accident Investigation Branch, Reports and bulletins, http://www.raib.gov.uk/publications/investigation_reports.cfm.
- [14] A. Kawaguchi, M. Miwa, and K. Terada, "Actual data analysis of alignment irregularity growth and its prediction model," *Quarterly Report of RTRI*, vol. 46, no. 4, pp. 262–268, 2005.
- [15] R. M. Alfelou, G. A. Carr, and M. Fateh, "Track degradation assessment using gage restraint measurements," *Transportation Research Record*, no. 1742, pp. 68–77, 2001.
- [16] C. Meier-Hirmer, G. Riboulet, F. Sourget, and M. Roussignol, "Maintenance optimization for a system with a gamma deterioration process and intervention delay: application to track maintenance," *Journal of Risk and Reliability*, vol. 223, no. 3, pp. 189–198, 2009.

- [17] P. Xu, R. K. Liu, F. T. Wang, Q. X. Sun, and H. L. Teng, "A novel description method for track irregularity evolution," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1358–1366, 2012.
- [18] P. Xu, Q. X. Sun, R. K. Liu, and F. T. Wang, "A short-range prediction model for track quality index," *Journal of Rail and Rapid Transit*, vol. 225, no. 3, pp. 277–285, 2011.
- [19] R. Liu, P. Xu, and F. Wang, "Research on a short-range prediction model for track irregularity over small track lengths," *Journal of Transportation Engineering*, vol. 136, no. 12, pp. 1085–1091, 2010.
- [20] E. T. Sellig, G. M. Cardillo, E. Stephens, and A. Smith, "Analyzing and forecasting railway data using linear data analysis," in *Proceedings of the 11th International Conference on Computer System Design and Operation in Railways and other Transit Systems*, Toledo, Spain, September 2008.
- [21] D. K. O'Brien, *Optimal estimation and rail tracking analysis [dissertation]*, University of Massachusetts Lowell, Lowell, Mass, USA, 2005.
- [22] P. Xu, Q. X. Sun, R. K. Liu, and F. T. Wang, "Key equipment identification model for correcting milepost errors of track geometry data from track inspection car," *Transportation Research C*. In press.

Research Article

Study on Optimization of Night Illumination in Expressway Long Tunnels

Weihua Zhao, Hao Chen, Qiang Yu, and Haoxue Liu

School of Automobile, Chang'an University, Xi'an 710064, China

Correspondence should be addressed to Weihua Zhao, sdzwh@chd.edu.cn

Received 11 September 2012; Revised 6 November 2012; Accepted 21 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Weihua Zhao 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.

Night illumination of expressway long tunnels in China was far from drivers' demand and caused lots of traffic accidents and electricity wasting. In order to solve this problem, reasonable night luminosity in expressway long tunnels in China was studied. Based on a visual disturbance generation mechanism, the mistakes in night illumination of Chinese expressway long tunnels were analyzed. Real road experiments were conducted and fuzzy assessments were used to establish models. The results showed that night luminosity was too high and should be controlled below 9 Lx. In the entrance segment, the luminosity should be gradually increased from 0 Lx to 9 Lx. In the middle segment, the luminosity should be remained at 9 Lx. In the exit segment, the luminosity should be gradually decreased to 0 Lx. The optimized luminosity is highly consistent with the actual demand of drivers. Traffic safety was improved and electricity consumption was reduced.

1. Introduction

The environment inside and outside tunnels was of great difference, in particular the luminosity. Significant contrast of dark and bright increased the visual and psychological burden of drivers. In the daytime, the drivers would suffer dark adaptation when entering tunnels and light adaptation when leaving tunnels. While at night, drivers would suffer light adaptation when entering these tunnels and dark adaptation when leaving. All expressway long tunnels in China are equipped with lights. The ventilation and lighting specification of PRC only required luminosity in day [1]. The same illumination program were generally adopted in Chinese practices for both day and night. This illumination method worsen visual disturbance in night and increased electricity consumption. For example the expense of illumination electricity in Chinese Zhongnan mountain tunnel was higher than charge. The night luminosity in expressway long tunnel was too higher to meet the drivers' demand.

2. Literature Review

The Luminosity difference inside tunnel and outside was a major factor to traffic safety. At the same time, illumination consumption of large amounts of electricity was obvious. Many studies have been conducted about this problem.

The ventilation and lighting specification of PRC stipulated detailed parameters of luminosity of tunnels in daytime [1]. The specification ignored the fact that drivers would suffer different visual disturbances between day and night. It stipulated luminosity mainly based on the visual disturbance that drivers would suffer in tunnel in daytime. Environmental factors affected driving behavior and information perception [2, 3]. Although the exact luminosity maybe different in every country, the major factors considered were the same. The different country had built self-specification [1, 4]. International Commission on Illumination published "Guide for the Lighting of Road Tunnels and Underpasses" in 2004 [5]. Schreuder and Swart studied the energy saving in tunnel entrance lighting in order to decrease the electricity consumption [6]. Piarc also studied the problem of operational cost reduction of road tunnels and proposed the illumination solution [7]. In China, Xia analyzed some problems about tunnel illumination unsolved in the specification. He proposed that the problem should be studied from driver light and dark adaption but no luminosity value regarded as appropriate [8]. Qiu et al. studied the specification and compared it with CIE then proposed the dynamic requiring illumination in the transition section of the tunnel lighting [9]. Tu and Chen Studied the illumination parameter of short tunnels [10]. Zhang aims at medium and short tunnels in expressway studied illumination [11]. Tunnel entrance and exit were regarded as black points of traffic accident so, Wang et al., analyze the illumination design at tunnel entrance and exit [12]. Schreuder researched the illumination in tunnel entrances at day [13]. Zhao and Liu from Chang'an University, investigated drivers' visual feature variation pattern and driving behaviors in tunnels and established relevant models to describe the variation pattern [14]. Fukuda selected one tunnel to study the eye fixation point [15]. Wang et al. collected data of Chinese traffic accidents happened in tunnel and proposed the high rate segment in tunnel [16]. Du et al. from Tongji University evaluated driving safety level in tunnels using drivers' pupil area variation and put forward many conclusions. From the point of content and methods, this study attributed pupil area variation completely to drivers' nervousness and ignored the impact of illumination and dark adaptation time. When driving in tunnel the relation between visual information perception and safety was analyzed [17, 18]. Schreuder, describes the relation between visual performance and road safety [19]. In china tunnel operation difficulties was researched including illumination [20]. In recent two years researchers in China mainly concerned in LED applied in tunnel [21]. All these studies were connected to the illumination in expressway tunnel but the conclusions could not be applied directly to the Chinese expressway long tunnel.

Although some studies on tunnel lighting were conducted, Chinese expressway long tunnel illumination at night is unreasonable yet. The reason for the above situation was that the illumination design ignored the sequence difference of dark and light adaptation between day and night, did not evaluate environment consistency, and only considered that illumination in tunnels could improve visual field. In contrast, the calculation method of illumination abroad was of great difference from the demand of drivers in China due to differences of vehicle velocity, traffic condition, and human factors.

3. Research Method and Experiment

3.1. Research Method

It was difficult to find out the exact luminosity. The reason was that human factors were not easy to quantitatively describe. Whether the luminosity was satisfying to drivers' demand only depended on the drivers' assessment. But the luminosity that could be applied in experiments was restricted to real tunnel illumination. If this problem was researched in laboratory, the drivers could not obtain the same feeling from the stimulated environment compared with real driving environment.

To solve this significant problem, this paper would conduct the real road experiments. Drivers passed through the tunnels with a different luminosity and assessed whether the illumination was appropriate. Considering the fact that people could not give exact assessment results, fuzzy mathematics would be the best method to calculate the luminosity. Aiming at the calculated luminosity, some drivers were selected to verify whether the optimized luminosity was right. Via the above methods, the best illumination parameters would be certain. Based on these analyses, this paper researched appropriate night luminosity in expressway long tunnels fit for Chinese real character of traffic condition and human factors and proposed an optimized night lighting solution.

3.2. Experiments

3.2.1. Subjects

As the uniform, requirements for more than 2 years of expressway driving experience were required. Also, the drivers should have good driving skills and normal visual ability, without physiological defect, or serious or severe accident experience. The subjects were from different industries with different experiences. They had different occupations and years of driving. At the same time, age and gender were considered. At last, including 12 female drivers total, 37 drivers were selected in the experiments.

3.2.2. Instruments

Luminosity meter was used to measure luminosity on ground and wall in tunnels. A 5-seat-car was chosen as an experiment vehicle. The vehicle velocity variation was recorded by Microwave radar speed sensor.

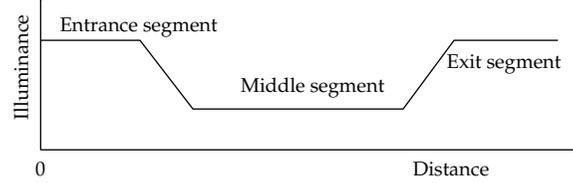
3.2.3. Tunnels and Luminosity

Before they entered each experimental tunnel, the drivers were required to run on an expressway without light for more than 30 minutes to make sure of full dark adaptation to the luminosity outside the tunnels. All the tunnels were one-way traffic and without interference from vehicles in the other way. Other vehicles would not pose any impact on the ambient luminosity. The length and luminosity of each tunnel were shown in Table 1.

All the tunnels had the same night and day illumination. The luminosity in middle segment was lower than both ends. The illumination was enhanced in entrance and exit segment. The luminosity variation pattern could be seen in Figure 1.

Table 1: Tunnels length and luminosity in middle segment.

Name of tunnel	Yongxing	Lijiahe	Nanwutai	Zhongnanshan	Qingcha
Length/m	1140	3650	2561	18020	1343
Illumination/Lx	60	50	40	30	20

**Figure 1:** Luminosity variation pattern.**Table 2:** Membership grade between luminosity and fuzzy assessment.

Luminosity	20	30	40	50	60
Appropriate	0.20	0.10	0.00	0.00	0.00
Less appropriate	1.00	0.31	0.00	0.00	0.00
All right	0.64	1.00	0.37	0.25	0.20
Brighter	0.12	0.73	1.00	0.90	0.85
Too bright	0.00	0.00	0.90	0.95	1.00

3.3. Experiment Program

The drivers did not know the goal of experiment and other factors were controlled as much as possible. The drivers entered the tunnels with their own driving habits after they had full adaptation outside the tunnels on the expressway. The drivers made subjective assessment of the appropriateness of luminosity in course of driving. Assessment results were recorded by the examiner.

4. Calculation and Analysis

4.1. Luminosity in Middle Segment

To ease dark and light adaptation of drivers, transition of tunnel luminosity should be smooth and comfortable. The luminosity in entrance segment should be gradually increased to the luminosity value in the middle segment, then retained at an appropriate value in middle segment, and then gradually decreased to 0 Lx in the exit segment. Therefore, the appropriate luminosity value in middle segment of expressway long tunnel was a critical parameter for illumination design.

4.1.1. Fuzzy Assessment

According to the illumination of the experimental tunnels, the domain of luminosity value in middle segment was taken as $L_m = \{60, 50, 40, 30, 20\}$. To simplify the assessment model, only the luminosity value in middle segment was assumed to be an assessment factor. Cognitive

fuzzy set of luminosity value in middle segment of expressway long tunnel at night was established. Fuzzy set of luminosity appropriateness in middle segment: $D = \{\text{too bright, lighter, all right, less appropriate, appropriate}\}$ and recorded as $\{D_1, D_2, D_3, D_4, D_5\}$. The membership grade of each was obtained using a fuzzy statistical method and was recorded as R_m . By taking all the results into consideration, the fuzzy relation between the luminosity value and appropriateness was obtained, as shown in Table 2.

Effective luminosity data of expressway long tunnel at night was obtained through luminosity assessment experiment. Using summary of assessment of tunnel luminosity cognition, membership grade of each fuzzy set of luminosity appropriateness was obtained. Membership grade variation showed that with the gradual decrease of luminosity value, the luminosity in middle segment was closer to the actual demand of drivers. This showed that the luminosity at night in middle segment of the Chinese expressway tunnels is too high.

4.1.2. Model and Results

Table 2 showed that the luminosity value in middle segment was smaller the membership grade of subset {appropriate} against the fuzzy set was higher. Considering road ambient consistency the luminosity in middle segment was smaller the ambient consistency, was higher. According to Weber's Law, the higher the illumination was, the smaller, equivalent variation of physical quantity that drivers can feel. On the basis of the above analysis and in combination of fuzzy mathematical membership function that fuzzy distribution of appropriate night luminosity in expressway long tunnels was "small" was certain. So that normal distribution was fit for the luminosity optimization function model. The specific calculation was shown in the following formula:

$$A(x) = \begin{cases} 1, & x \leq a \\ e^{-((x-a)/\sigma)^2}, & x > a. \end{cases} \quad (4.1)$$

where A is membership grade function, σ is standard deviation, and a is optimal illumination.

The fuzzy membership grade values in Table 2 defined by experiments were incorporated into function then the unknown number in formula (4.1) can be solved as optimal luminosity value. The specific equation was shown in the following formula:

$$\begin{aligned} e^{-((30-a)/\sigma)^2} &= 0.1 \\ e^{-((20-a)/\sigma)^2} &= 0.2. \end{aligned} \quad (4.2)$$

The calculation results were $a = 9$ and $\sigma = 2$. So that luminosity in middle segment was less than 9 Lx was appropriate for actual demand of drivers.

4.1.3. Result Validation

To validate the reliability of result drivers' fuzzy evaluation experiment was performed. 23 drivers were randomly selected from all subjects to participate in this experiment. The luminosity was set at 10, 8, 6, 4, 2, and 0 Lx, respectively. The results showed that when

Table 3: Luminosity in entrance segment.

Distance away from the entrance	0	30	50	100	150	200
Illumination/Lx	0	2	5	5	7	7

Table 4: Assessment results of easing lighting entrance segment.

Comment	Too bright	Brighter	Appropriate	Darker	Too dark
New entrance illumination	0.00	0.00	0.90	0.5	0.5
Original entrance illumination	0.8	0.2	0.0	0.0	0.0

the luminosity in the middle segment was less than 9 Lx membership grades of subjective luminosity {appropriate} levels given by drivers were all more than 0.95. This indicated that the calculation result was effective.

4.2. Luminosity in Entrance

When the night ambient luminosity outside the tunnel was 0 Lx, drivers would inevitably suffer light adaptation at entrance if illumination devices are turned on. This destroyed ambient consistency of road. So the ambient luminosity should be gradually increased to 9 Lx.

4.2.1. Optimization of Entrance Illumination

The light entered drivers eyes was ecological light reflected by every surface. The tunnel had been several reflection surfaces and was helpful to percept information. Due to diffuse reflection in tunnels, the light issued by illumination devices was more than a spot, but a domain. Meanwhile, all lamps in practices had been a certain power restriction. In consideration of convenience and feasibility, a simple solution was more reasonable. The result was shown in Table 3.

4.2.2. Illumination Assessment in Entrance

To verify whether this solution was effective, a comparison experiment of fuzzy assessment was performed. Randomly selected 21 drivers from all subjects were chosen to participate in this experiment. The experiment design was the same as to experiment on luminosity in middle segment. Domain $L_e = \{\text{original entrance illumination, new entrance illumination}\}$. Fuzzy set of appropriate luminosity in entrance segment with easing illumination was $E = \{\text{too bright, brighter, all right, darker, too dark}\}$, and recorded as $\{E_1, E_2, E_3, E_4, E_5\}$. The membership grade of each was obtained using a fuzzy statistical method, and was recorded as R_e . By taking all results into consideration, the fuzzy relation between the luminosity and its appropriateness in entrance segment with easing lighting was obtained, as shown in Table 4.

The experiment results showed that too high luminosity value in the original entrance segment might easily result in sudden environment change. Comparatively, the new illumination in entrance segment could meet the demand of drivers.

Table 5: Luminosity in entrance exit segment.

Distance away from the exit	0	30	50	100	150	200	250
Illumination/Lx	0	2	2	5	5	7	7

Table 6: Assessment results of lighting easing in exit segment.

Comment	Too light	brighter	appropriate	darker	Too dark
New exit illumination	0.00	0.00	0.95	0.5	0.0
Original exit illumination	0.9	0.1	0.0	0.0	0.0

4.3. Luminosity in Exit Segment

When drivers left the tunnel at night they would suffer dark adaptation. The luminosity at the exit that was higher dark adaptation problem would be more serious.

4.3.1. Optimization of Exit Illumination

The dark adaptation took longer time than light adaptation. On the contrary to entrance segment with lighting easing the optimal illumination in exit segment should be decreased gradually from 9 Lx to 0 Lx. The result was shown in Table 5.

4.3.2. Illumination Assessment in Exit

Fuzzy assessment of exit segment with easing lighting was performed in the same way with that of entrance segment. The fuzzy relation between the luminosity and its appropriateness in exit segment with easing lighting was obtained, as shown in Table 6.

The results showed that luminosity value in the original exit illumination was too high. The new exit illumination proposed based on appropriate luminosity value in the middle segment could better meet the demand of drivers.

5. Conclusions

Measurements of luminosity in more than 100 tunnels in Shanxi and Guangdong Provinces in China showed that the luminosity in middle segment was more than 20 Lx and higher at both ends for all tunnels. But the drivers feel that luminosity value was set as 9 Lx was appropriate in middle segment. The luminosity should be gradually increased from 0 Lx in entrance segment, maintained constant in middle segment, and then reduced to 0 Lx gradually in exit segment. According to these research results, night illumination of expressway long tunnels should be optimized. The optimization illumination could remove dark and light adaptation caused by luminosity difference when driving through tunnels at night. This illumination design could reduce accidents induced by visual disturbance and electricity power consumption. The following conclusions are obtained.

- (1) The luminosity should be controlled less than 9 Lx in the middle segment of expressway long tunnels at night. This illumination level was less than the level required in specification.

- (2) In the entrance segment the luminosity should be increased gradually from 0 Lx to 9 Lx.
- (3) The luminosity should be decreased gradually from 9 Lx to 0 Lx in the exit segment.

In this paper, the optimization of night illumination in expressway long tunnels was obtained according to Chinese traffic condition and character of drivers. The conclusion was verified by lighting practice in Yongxing tunnel and Wohuwan tunnel. In fact the luminosity of each point in expressway long tunnel could not be exactly equal to the proposed ones for optical line entered the drivers eye was reflected from driving environment. In practice, the luminosity in middle segment controlled under 9 Lx was the key problem. Such optimized night illumination in expressway long tunnel would not only eliminate the dark and light adaptation problem but also significantly reduce electricity power consumption.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51178054) the Fundamental Research Funds for the Central Universities (CHD2011JC031).

References

- [1] *JTJ 026.1-1999 Specifications for Design of Ventilation and Lighting of Highway Tunnel*, People's Education Press, Beijing, China, 2000.
- [2] W. H. Wang, W. Zhang, H. Bubb, and K. Ikeuchi, "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research Part C-Emerging Technologies*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [3] H. W. Guo, W. H. Wang, W. W. Guo, X. B. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [4] "Norwegian Public Road Administration Road Tunnels," Norwegian, NPRA Printing Center, 2004.
- [5] CIE 88:2004, *Guide for the Lighting of Road Tunnels and Underpasses*, International Commission on Illumination, Vienna, Austria, 2004.
- [6] D. A. Schreuder and L. Swart, "Energy Saving in Tunnel Entrance Lighting," 1993, <http://wenku.baidu.com/view/46cf968ccc22bcd126ff0c2d.html>.
- [7] Piarc, *Reduction of Operational Cost of Road Tunnels*, Piarc, 1999.
- [8] Y. X. Xia, "Problems and countermeasures of highway tunnel lighting," *Highway Tunnel*, no. 3, pp. 55–57, 2009.
- [9] F. Qiu, X. J. Ma, N. T. Liu, W. Lin, and X. Z. Liu, "Discussion on the dynamic requiring luminance in the transition section of the tunnel lighting," *Zhaoming Gongcheng Xuebao*, vol. 21, no. 6, pp. 13–18, 2010.
- [10] G. Tu and J. Z. Chen, "Study on illumination parameter of short tunnels," *Journal of Highway and Transportation Research and Development*, no. 6, pp. 125–128, 2009.
- [11] Y. L. Zhang, *Research on illumination of medium and short tunnels in expressway [M.S. thesis]*, Hunan University, Changsha, China, 2008.
- [12] Y. Wang, Z. Y. Guo, and Z. G. Liao, "Safety analysis for illumination design at tunnel entrance and exit," in *Proceedings of the International Conference on Intelligent Computation Technology and Automation (ICICTA '10)*, vol. 3, pp. 255–259, May 2010.
- [13] D. A. Schreuder, "Predetermination of the luminance in tunnel entrances at day," Tech. Rep., Institute for Road Safety Research SWOV, Leidschendam-Voorburg, The Netherlands, 1988.
- [14] W. H. Zhao and H. X. Liu, "Drivers' visual feature variation in long-tunnel exit of expressway," in *Proceeding of the First International Conference on Transportation (ICTIS '11)*, pp. 42–52, June 2011.
- [15] R. Fukuda, "Experimental considerations on the definition of eye fixation points," *Journal of Japan Ergonomics Society*, vol. 28, no. 2, p. 197, 1996.
- [16] H. Wang, H. X. Liu, W. H. Zhao et al., "Analysis on traffic accident feature in expressway tunnel," *Highway*, no. 11, pp. 144–147, 2009.

- [17] Z. G. Du, X. D. Pan, and X. B. Guo, "Experimental studies of visual adaptation during freeway tunnel's entrance and exit," *Journal of Harbin Institute of Technology*, vol. 39, no. 12, pp. 1998–2001, 2007.
- [18] Z. G. Du, X. D. Pan, and X. B. Guo, "Evaluation index's application studies on safety at highway tunnel's entrance and exit," *Journal of Tongji Institute of Technology*, vol. 36, no. 3, pp. 325–329, 2008.
- [19] D. A. Schreuder, "Visual performance and road safety," Tech. Rep., Institute for Road Safety Research SWOV, Leidschendam-Voorburg, The Netherlands, 1988.
- [20] Z. Y. Guo, L. Q. Kong, and Z. Yang, "Investigation report of operating safety research," Tech. Rep., Institute for Road Safety Research SWOV, Leidschendam-Voorburg, The Netherlands, 1988.
- [21] C. Y. Yang, Y. K. Hu, and Z. L. Chen, "Study on road lighting energy saving based on mesopic theory," *China Illuminating Engineering Journal*, vol. 19, no. 4, pp. 44–47, 2008.

Research Article

Pedestrian Walking Behavior Revealed through a Random Walk Model

Hui Xiong, Liya Yao, Huachun Tan, and Wuhong Wang

Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100084, China

Correspondence should be addressed to Hui Xiong, xionghui@bit.edu.cn

Received 27 September 2012; Accepted 4 November 2012

Academic Editor: Geert Wets

Copyright © 2012 Hui Xiong 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 applies method of continuous-time random walks for pedestrian flow simulation. In the model, pedestrians can walk forward or backward and turn left or right if there is no block. Velocities of pedestrian flow moving forward or diffusing are dominated by coefficients. The waiting time preceding each jump is assumed to follow an exponential distribution. To solve the model, a second-order two-dimensional partial differential equation, a high-order compact scheme with the alternating direction implicit method, is employed. In the numerical experiments, the walking domain of the first one is two-dimensional with two entrances and one exit, and that of the second one is two-dimensional with one entrance and one exit. The flows in both scenarios are one way. Numerical results show that the model can be used for pedestrian flow simulation.

1. Introduction

In recent years, modeling pedestrian flow has attracted considerable attention, partly because the model serves as basis for efficient crowd evacuation management and pedestrian facility operations. However, the research is still in its infancy owing to the complexity of human being's behaviors.

Most of the existing models for pedestrian flow are of microscopic nature, describing in detail the interactions among pedestrians, and between pedestrians and obstacles. Those models include, among others, cellular automata models [1–7], lattice gas models [8–12], the social force models [13], the centrifugal force models [14], and the floor field models [15, 16]. In cellular automata models, the walking space is two-dimensional and divided into cells. Each cell can either be empty, be occupied by exactly one pedestrian, or contain an obstacle. Cellular automata models are widely used for capturing pedestrian walking behaviors, such as bi-direction movement [1, 3, 4], pedestrian counter flow with different walk velocities

[2] or with right-moving preference [5], freezing transition phenomenon [6], and moving pedestrians' reaction to an obstacle [7]. Like cellular automata models, the lattice gas model consists of a set of stochastic rules on the square lattice. Applying the lattice gas model, previous studies have systematically investigated the phenomenon of jamming transition from the moving state at a low density to the stopping state at a higher density [8–11]. Jiang and Wu employed the lattice gas model to examine the interaction between a large object and pedestrians in the narrow channel [12]. The social force model, proposed by Helbing and Molnár [13], consists of three force terms that are measures for the moving motivations of a pedestrian. The model is able to reproduce the self-organization of collective phenomena of walking behaviors. By considering the headway and relative velocity among pedestrians, the centrifugal force model is capable of describing the behavior of lane formation [14]. The floor field model is yet another type of cellular automata models extended by realistically taking into account pedestrians' behaviors around the exit [15, 16].

Macroscopic models are often in the form of partial differential equations. Instead of describing individual pedestrian's behavior, this type of models treats the crowd as a whole and applies the conservation laws to capture the relationship among speed, flow, and density of pedestrian flow [17–21]. Directly starting from the flow conservation, Hughes [17] derived partial differential equations for flows with single or multiple pedestrian types. Likewise, Colombo and Rosini [18] introduced another partial differential equation model for pedestrian flows with a new parameter called characteristic density, which is used to reveal the maximal density in panic. Henderson [19] considered the movement of a crowd as an analogous system of gas molecules and applied the Maxwell-Boltzmann theory to describe the velocity distribution of people movements. Without making use of the conservation assumptions in Henderson's study, Helbing [20] developed a fluid dynamic model for the collective movement of pedestrians based on the Boltzmann-like approach.

Motivated by the work by Barkai et al. [22], this paper attempts to apply the approach of continuous-time random walks (CTRW) to derive a partial differential equation model to describe the motion of pedestrians. The CTRW is a useful model from statistical physics, in which each random particle jump is preceded by a random waiting time. Mathematically, the CTRW is a random walk subordinated to a renewal process. Different from the traditional macroscopic models, the proposed model is capable of not only capturing the macroscopic characteristics of pedestrian flows, but also describing the interactions among pedestrians, and between pedestrians and obstacles in terms of parameters in the model.

This paper is organized as follows. Section 2 introduces the continuous-time random walks model, formulated as a partial differential equation. Section 3 presents a high-order compact (HOC) solution scheme with the alternating direction implicit (ADI) method, followed by a numerical example in Section 4. Finally, Section 5 concludes the paper.

2. Model

Consider a random walker on a bounded two-dimensional lattice with a domain $\Omega \in \mathbb{R}^2$. It is assumed that there is no correlation between steps. Let τ represent the time that the walker stays at a particular point before making a "jump." Hereinafter we simply call τ as waiting time at a point and assume the waiting time at all points are independent random variables with an identical probability density function $\psi(\tau)$. We further define $\Psi(N, t)$

as the probability of N jumps occurring during the time interval $(0, t)$. Consequently, the probability that there is no jump occurring during the time interval $(0, t)$ is

$$\Psi(0, t) = P(\tau > t) = 1 - P(\tau \leq t) = 1 - \int_0^t \varphi(\tau) d\tau. \quad (2.1)$$

Applying the Laplace transform to both sides of (2.1) leads to

$$\tilde{\Psi}(0, s) = \frac{1 - \tilde{\varphi}(s)}{s}. \quad (2.2)$$

Note that $\Psi(1, t)$ represents the probability that the first jump occurs at time t' and there are no further jumps after it until t . We thus have

$$\Psi(1, t) = \int_0^t \varphi(t') \Psi(0, t - t') dt' \stackrel{\text{def}}{=} \varphi * \Psi, \quad (2.3)$$

where $\Psi(0, t)$ and $\varphi(t)$ are simplified as Ψ and φ . Similarly, we obtain

$$\Psi(N, t) = \varphi * \varphi * \varphi * \dots * \Psi = \varphi^{*N} * \Psi. \quad (2.4)$$

The Laplace form of (2.4) is given by

$$\tilde{\Psi}(N, s) = \tilde{\varphi}^N(s) \frac{1 - \tilde{\varphi}(s)}{s}. \quad (2.5)$$

Let $P(x, y, t)$ be the probability of observing the walker at site (x, y) at time t , and let $p_N(x, y)$ denote the probability that the walker is at the position (x, y) after N jumps. We then have

$$P(x, y, t) = \sum_{N=0}^{\infty} p_N(x, y) \Psi(N, t). \quad (2.6)$$

The Laplace transformation of (2.6) yields

$$\tilde{P}(x, y, s) = \frac{1 - \tilde{\varphi}(s)}{s} \sum_{N=0}^{\infty} p_N(x, y) \tilde{\varphi}^N(s) = \frac{1 - \tilde{\varphi}(s)}{s} p_0(x, y) + \frac{1 - \tilde{\varphi}(s)}{s} \sum_{N=1}^{\infty} p_N(x, y) \tilde{\varphi}^N(s). \quad (2.7)$$

In the domain of Ω with the length of the lattice edge being h (as shown in Figure 1), the law of total probability implies that

$$p_N(x, y) = r_1 p_{N-1}(x - h, y) + r_2 p_{N-1}(x, y - h) + r_3 p_{N-1}(x + h, y) + r_4 p_{N-1}(x, y + h), \quad (2.8)$$

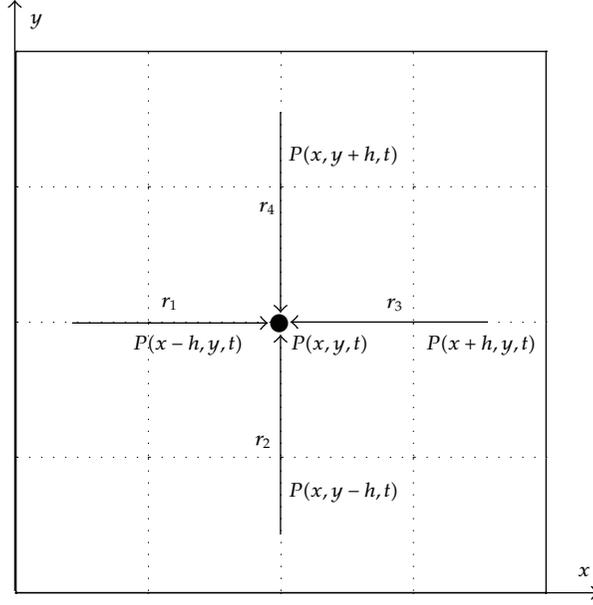


Figure 1: Probabilities of walking directions.

where r_1 to r_4 represents the probability for the walker to proceed forward, turn left, walk back, and turn right, respectively. The probabilities are assumed to be known and $\sum_{i=1}^4 r_i = 1$.

Applying Taylor's expansion to RHS of (2.8) at the point (x, y) leads to

$$\begin{aligned}
 p_N(x, y) &= r_1 \left[p_{N-1}(x, y) - h \frac{\partial}{\partial x_1} p_{N-1}(x, y) + \frac{1}{2} h^2 \frac{\partial^2}{\partial x_1^2} p_{N-1}(x, y) \right] \\
 &+ r_2 \left[p_{N-1}(x, y) - h \frac{\partial}{\partial x_2} p_{N-1}(x, y) + \frac{1}{2} h^2 \frac{\partial^2}{\partial x_2^2} p_{N-1}(x, y) \right] \\
 &+ r_3 \left[p_{N-1}(x, y) + h \frac{\partial}{\partial x_1} p_{N-1}(x, y) + \frac{1}{2} h^2 \frac{\partial^2}{\partial x_1^2} p_{N-1}(x, y) \right] \\
 &+ r_4 \left[p_{N-1}(x, y) + h \frac{\partial}{\partial x_2} p_{N-1}(x, y) + \frac{1}{2} h^2 \frac{\partial^2}{\partial x_2^2} p_{N-1}(x, y) \right] + O(h^2),
 \end{aligned} \tag{2.9}$$

where $O(h^2)$ denotes higher-order terms that are omitted hereinafter. We thus obtain the following:

$$\begin{aligned}
 p_N(x, y) &= p_{N-1}(x, y) - (r_1 - r_3) h \frac{\partial}{\partial x} p_{N-1}(x, y) - (r_2 - r_4) h \frac{\partial}{\partial y} p_{N-1}(x, y) \\
 &+ (r_1 + r_3) \frac{h^2}{2} \frac{\partial^2}{\partial x^2} p_{N-1}(x, y) + (r_2 + r_4) \frac{h^2}{2} \frac{\partial^2}{\partial y^2} p_{N-1}(x, y).
 \end{aligned} \tag{2.10}$$

Define $f(x) = (r_1 - r_3)/h$, $f(y) = (r_2 - r_4)/h$, where $f(x)$ is the direction force along the x -axis and $f(y)$ is the force along the y -axis. Consequently, (2.10) can be recast as

$$\begin{aligned} p_N(x, y) &= p_{N-1}(x, y) - V_1 \frac{\partial}{\partial x} p_{N-1}(x, y) - V_2 \frac{\partial}{\partial y} p_{N-1}(x, y) \\ &\quad + W_1 \frac{\partial^2}{\partial x^2} p_{N-1}(x, y) + W_2 \frac{\partial^2}{\partial y^2} p_{N-1}(x, y), \end{aligned} \quad (2.11)$$

where $V_1 = h^2 f(x)$, $V_2 = h^2 f(y)$, $W_1 = (r_1 + r_3)h^2/2$, and $W_2 = (r_2 + r_4)h^2/2$. Applying the Laplace transformation to (2.11), we have

$$\tilde{p}_N(x, y) = \tilde{p}_{N-1}(x, y) + h^2 L_{FP} \tilde{p}_{N-1}(x, y), \quad (2.12)$$

where $L_{FP} = -f(x)\partial/\partial x - f(y)\partial/\partial y + (W_1/h^2)\partial^2/\partial x^2 + (W_2/h^2)\partial^2/\partial y^2$.

Substituting (2.12) into (2.7) yields

$$\begin{aligned} \tilde{P}(x, y, s) &= \frac{1 - \tilde{\psi}(s)}{s} p_0(x, y) + \frac{1 - \tilde{\psi}(s)}{s} \tilde{\psi}(s) \sum_{N=0}^{\infty} [\tilde{p}_N(x, y) + h^2 L_{FP} \tilde{p}_N(x, y)] \tilde{\psi}^N(s) \\ &= \frac{1 - \tilde{\psi}(s)}{s} p_0(x, y) + \tilde{\psi}(s) \tilde{P}(x, y, s) + \tilde{\psi}(s) h^2 L_{FP} \tilde{P}(x, y, s). \end{aligned} \quad (2.13)$$

The specific form of (2.13) depends on the choice of the waiting time distribution $\psi(\tau)$. Several distributions are plausible based on the nature of pedestrian walking behaviors. Here we assume that $\psi(\tau)$ follows the exponential distribution, that is,

$$\psi(\tau) \sim \frac{1}{\lambda} e^{-\tau/\lambda}, \quad (2.14)$$

where λ is the mean of waiting time. The probability density function of waiting time in the Laplace space can be written as

$$\tilde{\psi}(s) = \frac{1}{1 + \lambda s} = 1 - \lambda s + \dots \quad (2.15)$$

Substituting (2.15) into (2.13) yields

$$\tilde{P}(x, y, s) = \lambda p_0(x, y) + (1 - \lambda s) \tilde{P}(x, y, s) + (1 - \lambda s) h^2 L_{FP} \tilde{P}(x, y, s). \quad (2.16)$$

Rearranging the above and dividing both sides by λs leads to

$$\tilde{P}(x, y, s) - \frac{1}{s} p_0(x, y) = \frac{k}{s} L_{FP} \tilde{P}(x, y, s), \quad (2.17)$$

where $k = \lim h^2/\lambda$ as $h^2 \rightarrow 0, \lambda \rightarrow 0$. Then the inverse Laplace transform of (2.17) becomes the following:

$$P(x, y, t) = kL_{FP} \int_0^t P(x, y, t) dt. \quad (2.18)$$

We thus obtain the continuous-time random walks model for pedestrian traffic as

$$\frac{\partial P(x, y, t)}{\partial t} = kL_{FP}P(x, y, t). \quad (2.19)$$

3. Numerical Algorithm

Equation (2.19) is recognized as the Fokker-Planck equation. If none of the coefficients is function of x or y , then it is a linear second-order partial differential equation. Here we employ a high-order compact scheme with the alternating direction implicit method to numerically solve this equation [23, 24]. Further denoting $(r_1 - r_3)k/h$ as β_x , $(r_2 - r_4)k/h$ as β_y , $(r_1 + r_3)k/2$ as α_x , and $(r_2 + r_4)k/2$ as α_y , we rewrite the model as follows:

$$\beta_x \frac{\partial P}{\partial x} + \beta_y \frac{\partial P}{\partial y} - \alpha_x \frac{\partial^2 P}{\partial x^2} - \alpha_y \frac{\partial^2 P}{\partial y^2} = -\frac{\partial P}{\partial t}. \quad (3.1)$$

Assume the domain Ω is divided evenly into spaced cells of length Δx along x -axis and length Δy along y -axis, and $\delta_x P_{ij}$, $\delta_y P_{ij}$, $\delta_x^2 P_{ij}$, and $\delta_y^2 P_{ij}$ represent the approximations to the first and second derivatives of P with respect to x or y at node (x_i, y_j) . Based on the standard central finite difference method, (3.1) can be discretized as follows:

$$\beta_x \delta_x P_{ij} + \beta_y \delta_y P_{ij} - \alpha_x \delta_x^2 P_{ij} - \alpha_y \delta_y^2 P_{ij} - \tau_{ij} = -\frac{\partial P}{\partial t} \Big|_{ij}. \quad (3.2)$$

In the above, the truncation error, that is, τ_{ij} , is

$$\tau_{ij} = \left[\beta_x \frac{\Delta x^2}{6} \delta_x^3 + \beta_y \frac{\Delta y^2}{6} \delta_y^3 - \alpha_x \frac{\Delta x^2}{12} \delta_x^4 - \alpha_y \frac{\Delta y^2}{12} \delta_y^4 \right] P_{ij} + O(\Delta x^4 + \Delta y^4), \quad (3.3)$$

where δ_x , δ_y , δ_x^2 , and δ_y^2 are the first- and second-order central difference operators.

Differentiating (3.1) with respect to x or y once and twice, respectively yields, approximations of higher-order derivatives as follows:

$$\begin{aligned}\delta_x^3 P_{ij} &= \left[\left(\frac{\beta_y}{\alpha_x} \delta_y - \frac{\alpha_y}{\alpha_x} \delta_y^2 \right) \delta_x + \frac{\beta_x}{\alpha_x} \delta_x^2 + \frac{1}{\alpha_x} \delta_x \delta_t \right] P_{ij} + O(\Delta x^4 + \Delta y^4), \\ \delta_y^3 P_{ij} &= \left[\left(\frac{\beta_x}{\alpha_y} \delta_x - \frac{\alpha_x}{\alpha_y} \delta_x^2 \right) \delta_y + \frac{\beta_y}{\alpha_y} \delta_y^2 + \frac{1}{\alpha_y} \delta_y \delta_t \right] P_{ij} + O(\Delta x^4 + \Delta y^4), \\ \delta_x^4 P_{ij} &= \left[\left(\frac{\beta_x \beta_y}{\alpha_x^2} \delta_y - \frac{\alpha_y \beta_x}{\alpha_x^2} \delta_y^2 \right) \delta_x + \left(\frac{\beta_x^2}{\alpha_x^2} + \frac{\beta_y}{\alpha_x} \delta_y - \frac{\alpha_y}{\alpha_x} \delta_y^2 \right) \delta_x^2 + \left(\frac{\beta_x}{\alpha_x^2} \delta_x + \frac{1}{\alpha_x} \delta_x^2 \right) \delta_t \right] P_{ij} \\ &\quad + O(\Delta x^4 + \Delta y^4), \\ \delta_y^4 P_{ij} &= \left[\left(\frac{\beta_x \beta_y}{\alpha_y^2} \delta_x - \frac{\alpha_x \beta_y}{\alpha_y^2} \delta_x^2 \right) \delta_y + \left(\frac{\beta_y^2}{\alpha_y^2} + \frac{\beta_x}{\alpha_y} \delta_x - \frac{\alpha_x}{\alpha_y} \delta_x^2 \right) \delta_y^2 + \left(\frac{\beta_y}{\alpha_y^2} \delta_y + \frac{1}{\alpha_y} \delta_y^2 \right) \delta_t \right] P_{ij} \\ &\quad + O(\Delta x^4 + \Delta y^4).\end{aligned}\tag{3.4}$$

Substituting (3.4) and (3.3) into (3.2) leads to

$$\left[\beta_x \delta_x + \beta_y \delta_y - A \delta_x^2 - B \delta_y^2 - C \delta_x \delta_y + D \delta_x \delta_y^2 + E \delta_x^2 \delta_y + F \delta_x^2 \delta_y^2 \right] P_{ij} = G \delta_t P_{ij} + O(\Delta x^4 + \Delta y^4),\tag{3.5}$$

where $A = \alpha_x - \beta_x^2 \Delta x^2 / 12 \alpha_x$, $B = \alpha_y - \beta_y^2 \Delta y^2 / 12 \alpha_y$, $C = \beta_x \beta_y \Delta x^2 / 12 \alpha_x + \beta_x \beta_y \Delta y^2 / 12 \alpha_y$, $D = \alpha_y \beta_x \Delta x^2 / 12 \alpha_x + \beta_x \Delta y^2 / 12$, $E = \alpha_x \beta_y \Delta y^2 / 12 \alpha_y + \Delta x^2 \beta_y / 12$, $F = \alpha_x \Delta y^2 / 12 - \alpha_y \Delta x^2 / 12$, and $G = -1 + (\beta_x \Delta x^2 / 12 \alpha_x) \delta_x + (\beta_y \Delta y^2 / 12 \alpha_y) \delta_y - (\Delta x^2 / 12) \delta_x^2 - (\Delta y^2 / 12) \delta_y^2$.

Following Karaa and Zhang [24], we define four finite difference operators,

$$\begin{aligned}L_x &= 1 + \frac{\Delta x^2}{12} \left(\delta_x^2 - \frac{\beta_x}{\alpha_x} \delta_x \right), & L_y &= 1 + \frac{\Delta y^2}{12} \left(\delta_y^2 - \frac{\beta_y}{\alpha_y} \delta_y \right), \\ A_x &= - \left(\alpha_x + \frac{\beta_x^2 \Delta x^2}{12 \alpha_x} \right) \delta_x^2 + \beta_x \delta_x, & A_y &= - \left(\alpha_y + \frac{\beta_y^2 \Delta y^2}{12 \alpha_y} \right) \delta_y^2 + \beta_y \delta_y.\end{aligned}\tag{3.6}$$

The difference between LHS of (3.5) and $(L_x A_y + L_y A_x) P_{ij}$ is expressed as

$$\left[\frac{\beta_x^2 \beta_y \Delta x^2 \Delta y^2}{144 \alpha_x \alpha_y} \delta_x^2 \delta_y + \frac{\beta_x \beta_y^2 \Delta x^2 \Delta y^2}{144 \alpha_x \alpha_y} \delta_x \delta_y^2 - \left(\frac{\beta_x^2 \Delta x^2 \Delta y^2}{144 \alpha_x} + \frac{\beta_y^2 \Delta x^2 \Delta y^2}{144 \alpha_y} \right) \delta_x^2 \delta_y^2 \right] P_{ij},\tag{3.7}$$

while the difference between RHS of (3.5) and $L_x L_y \delta_t P_{ij}$ is

$$\left[\frac{\beta_x \beta_y \Delta x^2 \Delta y^2}{144 \alpha_x \alpha_y} \delta_x \delta_y - \frac{\beta_x \Delta x^2 \Delta y^2}{144 \alpha_x} \delta_x \delta_y^2 - \frac{\beta_y \Delta x^2 \Delta y^2}{144 \alpha_y} \delta_x^2 \delta_y + \frac{\Delta x^2 \Delta y^2}{144} \delta_x^2 \delta_y^2 \right] \delta_t P_{ij}.\tag{3.8}$$

Thus (3.5) can be rewritten as

$$[L_x A_y + L_y A_x] P_{ij} = L_x L_y \delta_t P_{ij} + O(\Delta x^4 + \Delta y^4), \quad (3.9)$$

since adding the above two expressions to (3.5) will not influence the accuracy. Applying the Crank-Nicolson discretization, we have

$$(L_x A_y + L_y A_x) \frac{P_{ij}^{n+1} + P_{ij}^n}{2} = -L_x L_y \frac{P_{ij}^{n+1} - P_{ij}^n}{\Delta t} + O(\Delta x^4 + \Delta y^4) + O(\Delta t^2). \quad (3.10)$$

This discretization is apparently second order in time and fourth order in space.

We move the terms with P_{ij}^{n+1} in the above equation to its LHS and add $\Delta t^2 A_x A_y P_{ij}^{n+1}/4$ to it. Similarly, we move the terms with P_{ij}^n to the RHS and add $\Delta t^2 A_x A_y P_{ij}^n/4$. Consequently, (3.10) becomes the following equation after dropping the error terms [24]:

$$\left(L_x + \frac{\Delta t}{2} A_x\right) \left(L_y + \frac{\Delta t}{2} A_y\right) u_{ij}^{n+1} = \left(L_x - \frac{\Delta t}{2} A_x\right) \left(L_y - \frac{\Delta t}{2} A_y\right) P_{ij}^n. \quad (3.11)$$

Employing the alternating direction implicit method, we have

$$\begin{aligned} \left(L_x + \frac{\Delta t}{2} A_x\right) P_{ij}^{n+1/2} &= \left(L_x - \frac{\Delta t}{2} A_x\right) \left(L_y - \frac{\Delta t}{2} A_y\right) P_{ij}^n, \\ \left(L_y + \frac{\Delta t}{2} A_y\right) P_{ij}^{n+1} &= P_{ij}^{n+1/2}, \end{aligned} \quad (3.12)$$

where $P_{ij}^{n+1/2}$ is an intermediate variable.

4. Numerical Example

We applied the proposed model and solution algorithm to a rectangular walking platform with $x = 30$, $y = 20$. Pedestrians are only allowed to walk from the left boundary to the right one. There are two doors of the same width as $w = 4$ where pedestrians can enter the platform, as shown in Figure 2. Initially, there are no pedestrians in the platform. The boundary conditions are

$$P(x, 0, t) = 0, \quad P(x, 20, t) = 0, \quad P(0, [8, 12], t) = 0, \quad (4.1)$$

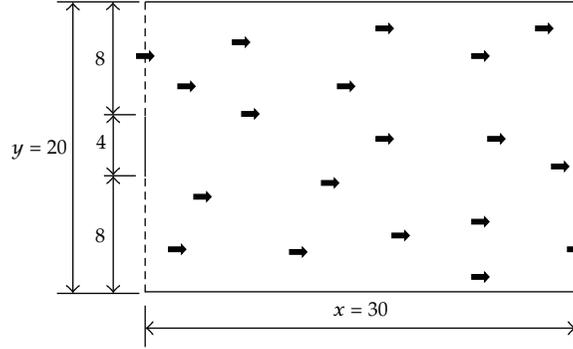


Figure 2: Schematic illustration of the one-way pedestrian flow in a platform. The length of the platform is 30 and the width is 20. The left dashed lines represent two entrances and the right one is exit. Solid lines are walls.

and the inputs are

$$\begin{aligned}
 P(0, d, t) &= \frac{t}{1.5}, & t \in [0, 1.5], \\
 P(0, d, t) &= 1.0, & t \in (1.5, 3.5), \\
 P(0, d, t) &= \frac{(5-t)}{1.5}, & t \in [3.5, 5], \\
 P(0, d, t) &= 0, & t > 5,
 \end{aligned} \tag{4.2}$$

where $d = (0, 8) \cup (12, 20)$.

We further assume r_1 to r_4 to be 0.70, 0.15, 0.0, and 0.15, respectively, and $\lambda = 0.045$ and $h = 0.05$.

Figure 3 is snapshots of numerical solutions of the pedestrian flow at times $t = 1, 2, \dots, 9$, respectively, to show the movement pattern of the pedestrians. The density increases steadily with the increase of entering flow and reaches its maximum at time $t = 5$. The density centered at either group is becoming less while the density between these two groups is becoming larger as the pedestrians are walking forward. For each group of pedestrians, density at the center is always larger than those in the surroundings. The reason is that the people around the block are much easier to disperse than those in the middle. As the time is close to $t = 9$, the density is approaching zero and only some late-entering or slow-walking people remain in the platform. From the snapshots, we can observe the phenomena of dispersion and advection of the pedestrian flow.

To further illustrate the model, additional experiments were conducted with the whole left boundary, that is, $x = 0$, as the entrance. In addition, the inflow is supposed to be steady with $P(0, y, t) = 1.0$, where $y \in (0, 20)$ and $t \in (0, 30]$. Figure 4 plots the density along x at $t = 1, 2, \dots, 8$, respectively. It can be observed that there is a sharp decrease in each curve, indicating that only a few pedestrians are fast walkers. The results are consistent with the phenomenon we may observe in reality that no matter how crowded a platform is, the density close to the exit is always less than the jam density.

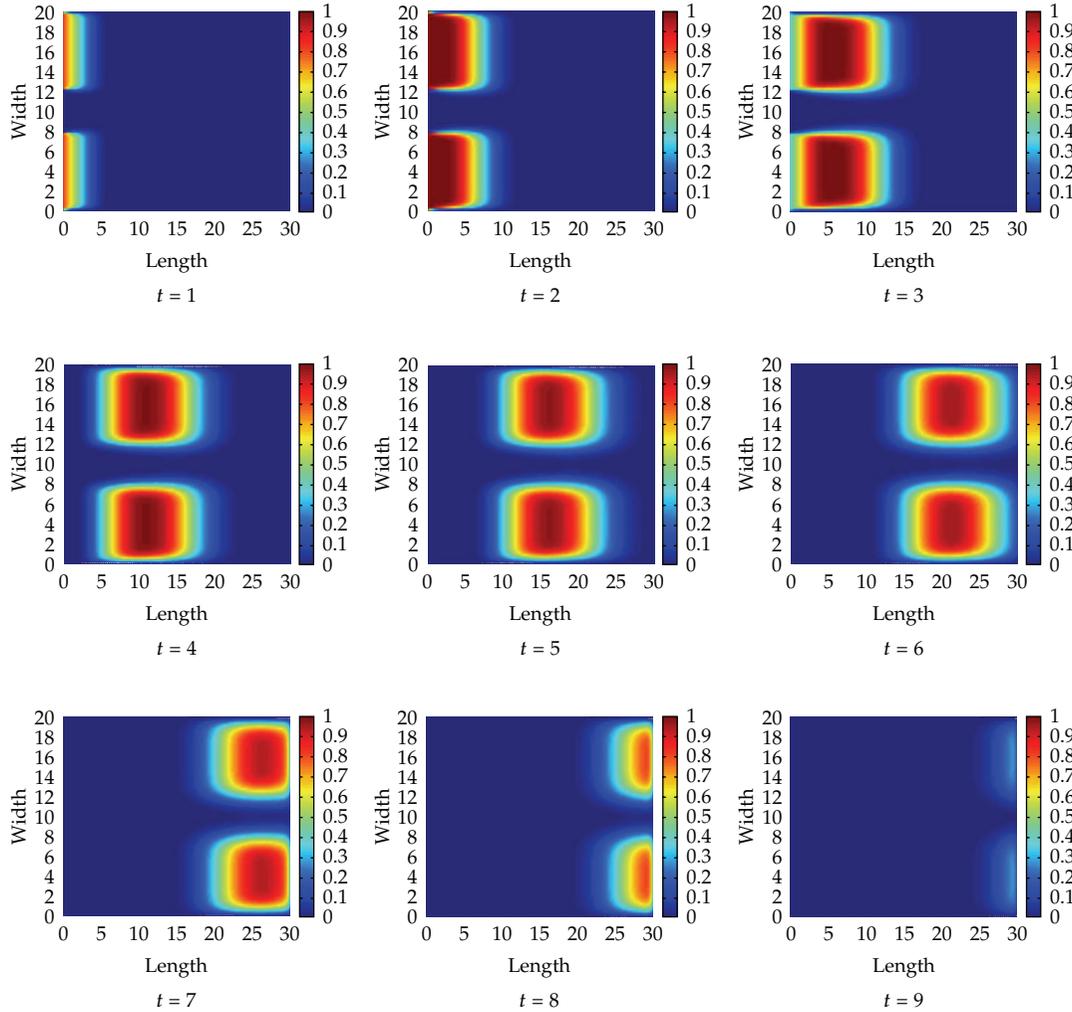


Figure 3: Density of pedestrian flow at different times.

To reveal the impact of direction choice behavior on the flow patterns, Figure 5 plots the average density of the platform along time under various scenarios where $r_1 = 0.80, 0.60, 0.40, 0.33$, respectively, and probabilities of left and right turns are $r_2 = r_4 = (1 - r_1)/2$ and walking backward is not allowed. It is shown that the time the flow becomes steady is significantly dependent on r_1 . The larger r_1 is, the faster the flow reaches to a steady state. The numerical results coincide with actual pedestrian moving behavior. Moreover, a smaller value of r_1 leads to a lower density in the steady state.

Figure 6 illustrates the average density across the time under different entering flow intensity at $PE = 1.0, 0.8, 0.6, 0.4, 0.2$, respectively. It shows the steady density is only slightly lower than the input flow intensity. In addition, it is observed that no matter what intensity the inflow is, the time the flow reaches to its steady state remains almost the same.

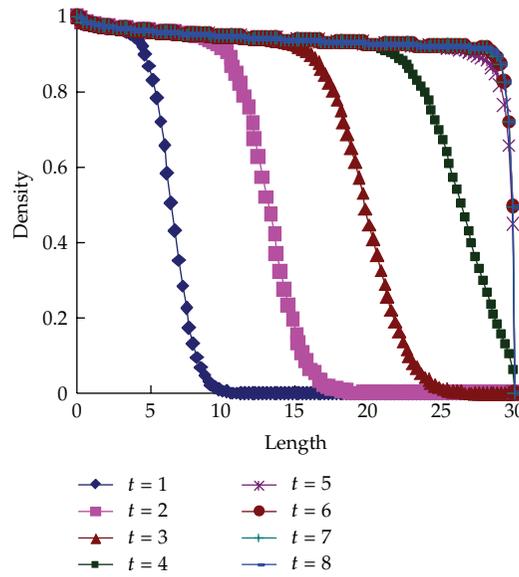


Figure 4: Density along x -axis at different times with $r_1 = 0.6$, $r_2 = 0.2$, $r_3 = 0.0$, and $r_4 = 0.2$.

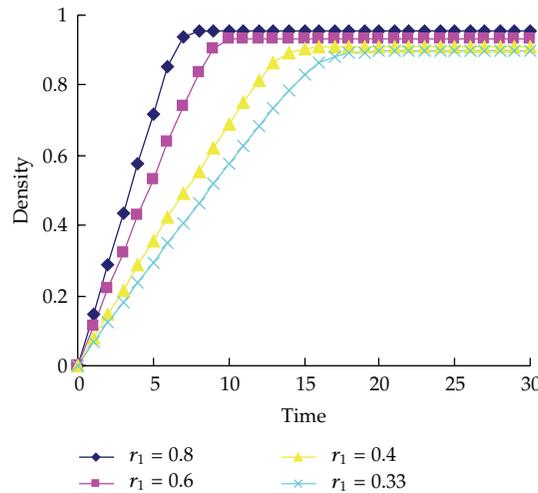


Figure 5: Average density of the platform along time with different r_1 's.

5. Conclusion

This paper is an application of continuous-time random walks approach to pedestrian flow simulation. The model is capable of describing macroscopic phenomena such as forward moving and dispersion of pedestrian flow. In addition, by varying coefficients of the model, some microscopic phenomena such as route/direction choice behaviors can be replicated. To solve the model, a high-order compact scheme with the alternating direction implicit method is applied. Numerical results validated both the model and the numerical method.

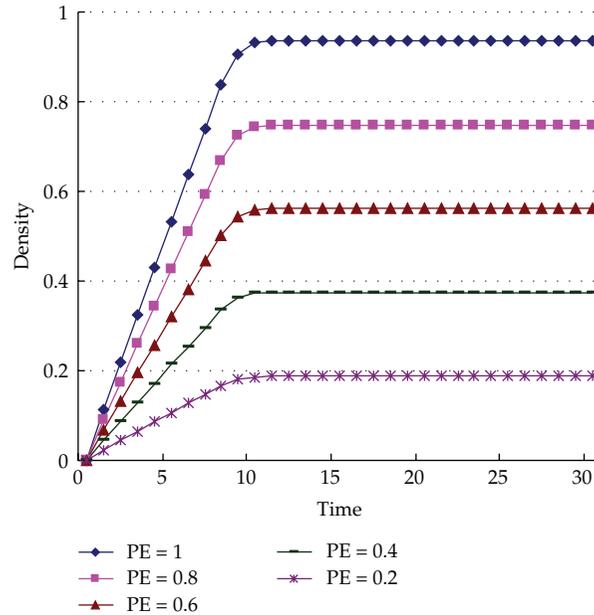


Figure 6: Average density with different entering flow intensity, where $r_1 = 0.6$, $r_2 = 0.2$, $r_3 = 0.0$, and $r_4 = 0.2$.

The model formulation in the paper only accounts for the distribution of the waiting time. In our future study, the probability distribution of jump length will be considered to further enhance the validity of the model. To make the model more practically applicable, we also plan to incorporate bi-direction flow and more realistic boundary conditions, such as input and output between platform and trains in platform.

Acknowledgment

This work is supported by Beijing Municipal Natural Science Foundation (Grant no. 8092026).

References

- [1] 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.
- [2] W. G. Weng, T. Chen, H. Y. Yuan, and W. C. Fan, "Cellular automaton simulation of pedestrian counter flow with different walk velocities," *Physical Review E*, vol. 74, no. 3, Article ID 036102, 7 pages, 2006.
- [3] V. J. Blue and J. L. Adler, "Cellular automata microsimulation for modeling bi-directional pedestrian walkways," *Transportation Research Part B*, vol. 35, no. 3, pp. 293–312, 2001.
- [4] Y. F. Yu and W. G. Song, "Cellular automaton simulation of pedestrian counter flow considering the surrounding environment," *Physical Review E*, vol. 75, no. 4, Article ID 046112, 8 pages, 2007.
- [5] L. Z. Yang, J. Li, and S. B. Liu, "Simulation of pedestrian counter-flow with right-moving preference," *Physica A*, vol. 387, no. 13, pp. 3281–3289, 2008.
- [6] T. Nagatani, "Freezing transition in bi-directional CA model for facing pedestrian traffic," *Physics Letters, Section A*, vol. 373, no. 33, pp. 2917–2921, 2009.

- [7] A. Varas, M. D. Cornejo, D. Mainemer et al., "Cellular automaton model for evacuation process with obstacles," *Physica A*, vol. 382, no. 2, pp. 631–642, 2007.
- [8] M. Muramatsu, T. Irie, and T. Nagatani, "Jamming transition in pedestrian counter flow," *Physica A*, vol. 267, no. 3, pp. 487–498, 1999.
- [9] M. Muramatsu and T. Nagatani, "Jamming transition of pedestrian traffic at a crossing with open boundaries," *Physica A*, vol. 286, no. 1, pp. 377–390, 2000.
- [10] Y. Tajima, K. Takimoto, and T. Nagatani, "Scaling of pedestrian channel flow with a bottleneck," *Physica A*, vol. 294, no. 1-2, pp. 257–268, 2001.
- [11] Y. Tajima and T. Nagatani, "Clogging transition of pedestrian flow in T-shaped channel," *Physica A*, vol. 303, no. 1-2, pp. 239–250, 2002.
- [12] R. Jiang and Q. S. Wu, "The moving behavior of a large object in the crowds in a narrow channel," *Physica A*, vol. 364, pp. 457–463, 2006.
- [13] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995.
- [14] W. J. Yu, R. Chen, L. Y. Dong, and S. Q. Dai, "Centrifugal force model for pedestrian dynamics," *Physical Review E*, vol. 72, no. 2, Article ID 026112, 7 pages, 2005.
- [15] D. Yanagisawa and K. Nishinari, "Mean-field theory for pedestrian outflow through an exit," *Physical Review E*, vol. 76, no. 6, Article ID 061117, 9 pages, 2007.
- [16] H. J. Huang and R. Y. Guo, "Static floor field and exit choice for pedestrian evacuation in rooms with internal obstacles and multiple exits," *Physical Review E*, vol. 78, no. 2, Article ID 021131, 9 pages, 2008.
- [17] R. L. Hughes, "A continuum theory for the flow of pedestrians," *Transportation Research Part B*, vol. 36, no. 6, pp. 507–535, 2002.
- [18] R. M. Colombo and M. D. Rosini, "Pedestrian flows and non-classical shocks," *Mathematical Methods in the Applied Sciences*, vol. 28, no. 13, pp. 1553–1567, 2005.
- [19] L. F. Henderson, "On the fluid mechanics of human crowd motion," *Transportation Research*, vol. 8, no. 6, pp. 509–515, 1974.
- [20] D. Helbing, "A fluid-dynamic model for the movement of pedestrians," *Complex Systems*, vol. 6, pp. 391–415, 1992.
- [21] Y. Xia, S. C. Wong, and C. W. Shu, "Dynamic continuum pedestrian flow model with memory effect," *Physical Review E*, vol. 79, no. 6, Article ID 066113, 8 pages, 2009.
- [22] E. Barkai, R. Metzler, and J. Klafter, "From continuous time random walks to the fractional Fokker-Planck equation," *Physical Review E*, vol. 61, no. 1, pp. 132–138, 2000.
- [23] W. F. Spitz and G. F. Carey, "High-order compact scheme for the steady stream-function vorticity equations," *International Journal for Numerical Methods in Engineering*, vol. 38, no. 20, pp. 3497–3512, 1995.
- [24] S. Karaa and J. Zhang, "High order ADI method for solving unsteady convection-diffusion problems," *Journal of Computational Physics*, vol. 198, no. 1, pp. 1–9, 2004.

Research Article

Analysis of the Travel Intent for Park and Ride Based on Perception

Huanmei Qin, Hongzhi Guan, and Guang Zhang

Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Huanmei Qin, hmqin@bjut.edu.cn

Received 10 September 2012; Revised 10 November 2012; Accepted 21 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Huanmei Qin 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 a multimodal travel behavior, park and ride includes several trip modes such as car, walking, bus, or railway. And people's choice of park and ride is influenced by many factors. This paper, based on the park and ride behavior survey in Beijing, will analyze the relationship between the perception of the influencing factors and the behavior intent for park and ride by using structural equation modeling. The conclusions suggest that the park and ride choice for travelers is a passive behavior which means giving up driving the car is mainly caused by the serious traffic congestion. Furthermore, improving the service level of the park and ride facilities and the comfort for riding bus or railway will increase the utilization of park and ride facilities. The perceptions of the influencing factors have both direct and indirect effects on the travel intent for park and ride by the interaction among the influencing factors.

1. Introduction

Park and ride is a travel mode, in which the car travelers who are going to the city center (or congestion area) park in the peripheral area and then reach the destination by using public transportation. This mode is helpful to reduce the car flow in the city center, relieve the traffic pressure; and perfect the urban traffic structure. The factors affecting the park and ride choice for the car drivers include personal characteristics, service level of the transfer facilities, and policy factors.

Hole has used stated choice (SC) data to forecast the demand for an employee park and ride service. Four different Logit models have been established in the paper. The model estimations show that the drivers with low income, more cars, and parking spaces scarcity at travel destination would be more likely to use park and ride [1].

Bos et al. have conducted a stated choice experiment in Nijmegen, The Netherlands, using the hierarchical information integration method. The results indicate that social safety, quality of the connecting public transport, and relative travel times by transport modes are

key attributes to the success of park and ride facilities. Contextual variables seem to have only a minor impact [2].

Hess has developed the multinomial mode Logit model including driving alone, riding in a carpool, or transferring to transit for the trip to work in Portland's CBD. The results suggest that raising the cost of parking at work sites and decreasing the transit travel time (by improving service and decreasing headways) will reduce the driving alone mode share [3].

At present, the construction of park and ride facilities is just beginning in metropolitan cities of China. For example, Beijing and Shanghai have launched pilot program of park and ride and made the plan of park and ride facilities. Some related researches are shown as follows.

Hui and He have studied the influencing factors and the travel intent for park and ride choice behavior and used disaggregated model to analyze the importance of influencing factors. The paper brings forward the idea that the main factors affecting park and ride choice behavior are passive factors such as road traffic congestion and lack of parking spaces [4]. He et al. have established a binary Logit model using the stated preference survey data in Nanjing, China. The results show that income, driving experience, trip purpose, traffic congestion level, and parking fee have significant impacts on the use of park and ride facilities [5].

The previous researches mainly analyze the relations between the influencing factors and choice behaviors by using the investigations and models. The travelers may have different perceptions of the same influencing factors, and then they will produce different travel intents under the same traffic condition [6]. It can be said that the perception of the influencing factors is an important aspect for travel behavior analysis.

Sakano and Benjamin have studied the travel behavior for choosing the commuter rail or car between two cities using the structural equation analysis. The model is specified and estimated to test various policy factors such as changes in fares, headway and travel time of train service, and the congestion on the highway. It was found that travel time and explicitly expected extra travel time due to congestion are not significantly related to mode choice, and primary activities are significant predictors of mode choice [7].

Lin et al. have proposed an alternative model specification for better conceptualizing the customer perceived value construct and to discuss the theoretical justification of the model. The proposed model was elaborated based on theoretical contexts. Three models of different conceptualization specifications were estimated and compared with eTail service value survey data [8].

Rahaman et al. have used SERVQUAL to measure the quality of provided services and analyze the gaps between customer expectations and perceptions of the service organization. The research points out how the management of service improvement can become more logical and integrated to prioritize service quality dimensions and affect on increasing or decreasing service quality gaps [9].

Yüksel has used a structural equation modeling to explore the relationships between the perceived shopping environment and tourists' emotions and shopping values. The emotional state and shopping value created by the shopping environment were found to influence the enjoyment of shopping, revisiting intentions, and the tendency to spend more money and time than originally planned [10].

This paper takes the car park and ride travel behavior as the research object. Based on the park and ride behavior survey conducted in Beijing, China, it will analyze the relations between the perception of various influencing factors and the behavior intent for park and

ride by using the structural equation modeling. The results will provide some references for the park and ride behavior analysis and facilities planning and location.

2. Park and Ride Behavior Survey

In order to analyze the relationship between the perceptive factors and the park and ride behavior intent, it is necessary to understand the perception level of the related influencing factors for travelers by the survey. The influencing factors of this survey include personal characteristics such as age, sex, monthly income, car ownership, and household size and the perception of the traffic congestion, parking space and parking fee at destination, parking and riding at park and ride facility, travel time, and travel cost as well as comfort for riding public transit. The perceptive questions are constructed in terms of a five-point scale that measures the level of satisfaction with a statement. The scale categories are identified as "strongly satisfied," "relatively satisfied," "generically satisfied," "unsatisfied" and "strongly unsatisfied." It can also change the expressed optional contents according to different questions. For example, the perception levels of the traffic condition are "very congested," "relatively congested," "congested," "not congested," and "smooth traffic flow".

This paper used the combined survey methods with revealed preference (RP) and stated preference (SP). RP refers to situations where travelers are observed to make choices in real situations and is used to investigate the travel time, travel cost, and weekly trip times by using park and ride for the travelers. SP refers to situations where choices are observed in the hypothetical situations and is used to investigate the intent for using park and ride in the future [11–13]. The survey respondents are travelers who possess cars and have used the park and ride facilities once. The survey place is Tiantongyuan North park and ride facility which is the terminal station for Beijing subway line 5 with the straight line distance from the city centre of about 20 km. In the survey, the investigator inquires the interviewees at designated time and place and makes them fill in the questionnaires and then retrieves them. The survey was conducted from February to March in 2011. The retrieved sample is 112, and the effective sample is 99.

For the sex distribution, the proportion of interviewees for male and female is 52% and 48%, respectively. For the age distribution, the age of most interviewees is between 20 to 50 years, and the proportion is 78%. For the monthly income distribution, the proportion of interviewees with the monthly income less than or equal to 3000 yuan is 49%. The proportion of those with the monthly income between 3000 and 5000 is 31%. For the weekly days distribution for using park and ride, the proportion of interviewees with one day for using park and ride is 14%. The proportion of those with two days and three days is 18% and 17%, respectively. The proportion of those with the four days and five days is 10% and 24%, respectively. This indicates commuters often use the park and ride.

3. Park and Ride Behavior Survey Analysis

According to the survey data, this section will analyze the influencing factors.

3.1. Analysis of Personal Characteristics

It is shown by Figures 1 and 2 that the proportion of people who possess one car is 82%, two cars is 16%. The people who possess three cars is relatively less, and the proportion is only

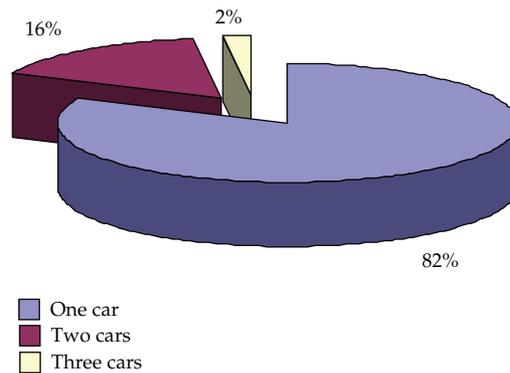


Figure 1: Car ownership distribution.

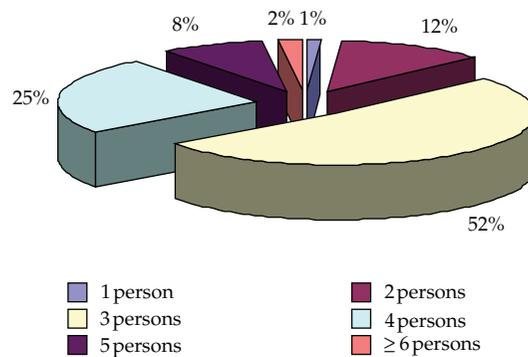


Figure 2: Household size distribution.

2%. The household size is mostly 2, 3, and 4 persons which altogether share the proportion of 89%.

3.2. Perception of the Traffic Condition

Figure 3 shows that the perception of traffic condition is relatively poor for the interviewees. The proportion of the persons who perceive traffic condition as relatively congested and very congested is 81%. The proportion of the interviewees who perceive traffic condition as not congested and smooth is only 3%. This indicates that the traffic condition for driving is poor.

As can be seen from Figure 4, 44% of interviewees feel that the parking fee at destination is a little higher, and 32% of interviewees feel that the parking fee is reasonable and acceptable. Only 5% of interviewees think that the parking fee is cheap.

3.3. Perception on the Service Level of Park and Ride Facilities

As shown in Figures 5 and 6, among the two main factors influencing the service level of the park and ride facility, the proportions of interviewees who consider generally and relatively satisfied with the services are 44% and 33%, respectively, and only 8% of interviewees

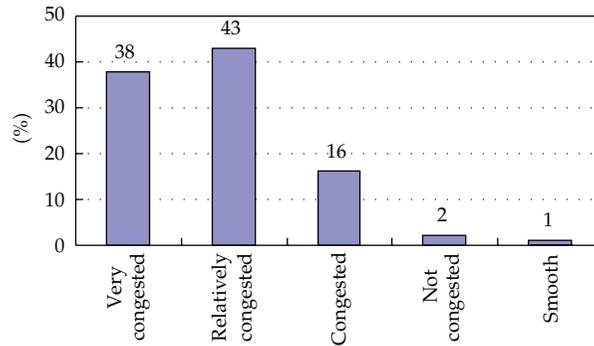


Figure 3: Traffic congestion.

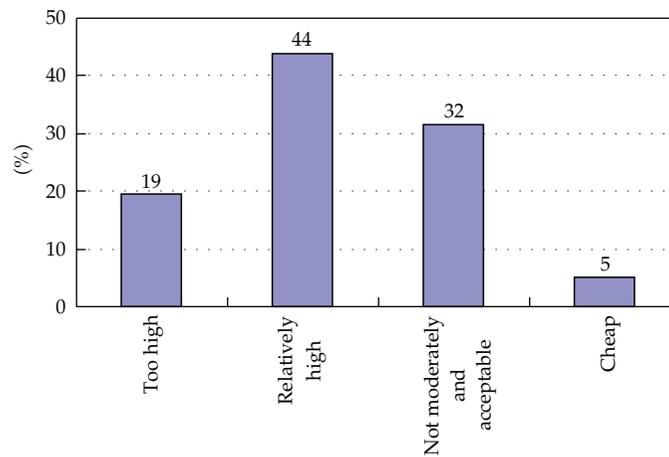


Figure 4: Parking fee at destination.

are unsatisfied or strongly unsatisfied. For transferring at the park and ride facility, the proportions of interviewees who consider generally and relatively satisfied are 40% and 41%, respectively, and only 6% of interviewees are unsatisfied or strongly unsatisfied. This indicates that the service level of park and ride facility is rather good and the travelers are quite satisfied with it.

3.4. Perception of Travel Time and Cost

As shown in Figure 7, 45% of park and ride users feel comparatively satisfied, and 36% of those feel quite satisfied for the travel time. Only 11% of those feel unsatisfied and strongly unsatisfied for the travel time. Figure 8 shows that 16% of the interviewees are very satisfied for the travel cost, and 51% of those are relatively satisfied for the travel cost. Only 6% of those are unsatisfied and strongly unsatisfied for the travel cost. This indicates that the park and ride users are satisfied for the travel time and cost.

The perception of the comfort for riding bus or railway is shown in Figure 9. 23% of interviewees express that it is too crowded to be tolerated. While 38% of those consider that it is relatively crowded and a little unbearable. 34% of them consider that it is crowded but

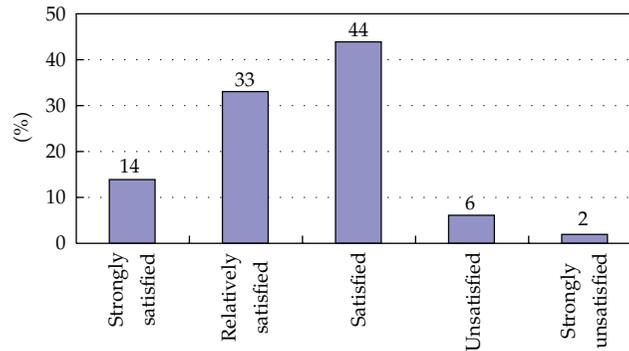


Figure 5: Satisfaction for parking at park and ride facility.

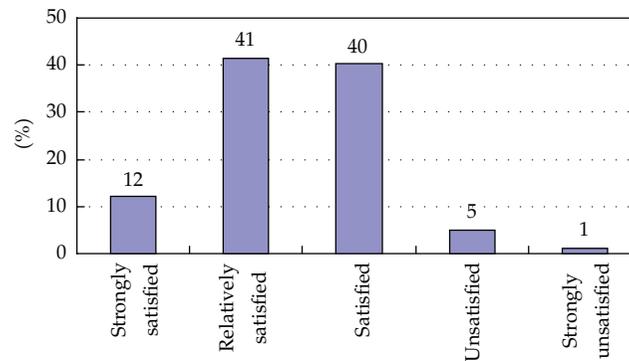


Figure 6: Satisfaction for transferring at park and ride facility.

acceptable. Only 5% of them consider that it is comfortable. This indicates that service level for public transit need, to be improved.

3.5. Park and Ride Travel Intent

Park and ride travel intent is represented by the change of weekly travel times for using park and ride in the future as shown in Figure 10. 42% of interviewees do not change their weekly travel times for using park and ride. 36% of interviewees will increase their travel times for using park and ride. Among this group of travelers, 7% increase park and ride by three times, 16% by twice, and 13% by once. By contrast, 22% of users like to reduce their frequency of park and ride.

4. Analysis of the Park and Ride Travel Intent

Based on the previous analysis for the survey data, the structural equation is used to analyze the relationship between the influencing factors and park and ride intent. As a method to estimate and verify the causal relationship among variables, the structural equation is mainly used for analyzing and simulating the latent variable that cannot be directly observed. Recently, structural equation is gradually applied to the travel behavior researches [14–17].

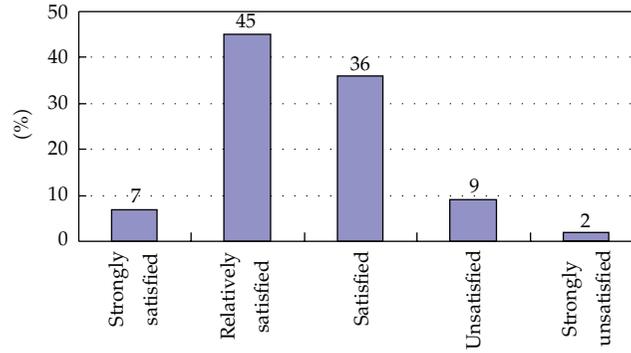


Figure 7: Satisfaction for travel time.

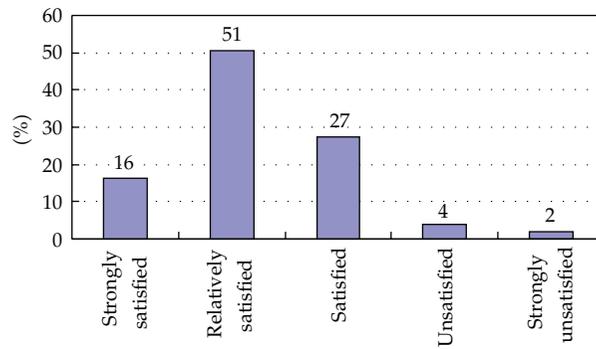


Figure 8: Satisfaction for travel cost.

4.1. Structural Equation Modeling Theory

Structural equation modeling (SEM) is a statistic analysis method of the variables' relation based on the covariance matrix. The steps for applying the structural equation modeling include the model specification, quantification of variables, identification, estimation, and modification [18, 19].

Structural equation modeling (SEM) is composed of measurement model and structural model. A general structural equation includes three formulas as follows. Formula (4.1) is the structural model which analyzes the relation among latent variables. formulas (4.2) and (4.3) are measurement models which analyze the relation between the latent variables and the observed variables:

$$\eta = B\eta + \Gamma\xi + \zeta, \tag{4.1}$$

$$y = \Lambda_y\eta + \varepsilon, \tag{4.2}$$

$$x = \Lambda_x\xi + \delta, \tag{4.3}$$

where η is the vector of endogenous latent variables. B is the coefficient matrix of endogenous latent variables; Γ is the coefficient matrix of exogenous latent variables; ξ is the vector of exogenous latent variables; ζ is the vector of residual error of the structural model; y is

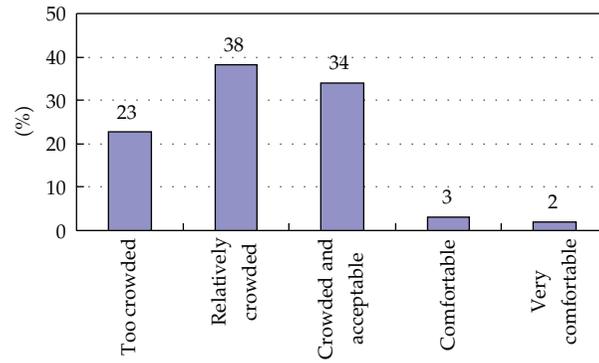


Figure 9: Perception of the comfort for riding bus or railway.

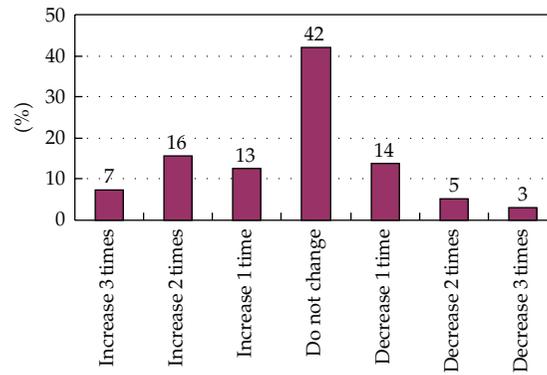


Figure 10: Park and ride travel intent.

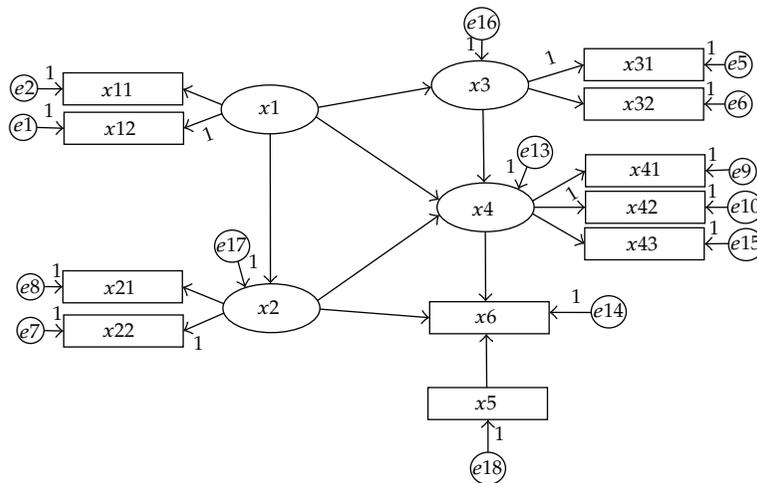


Figure 11: The model specification diagram.

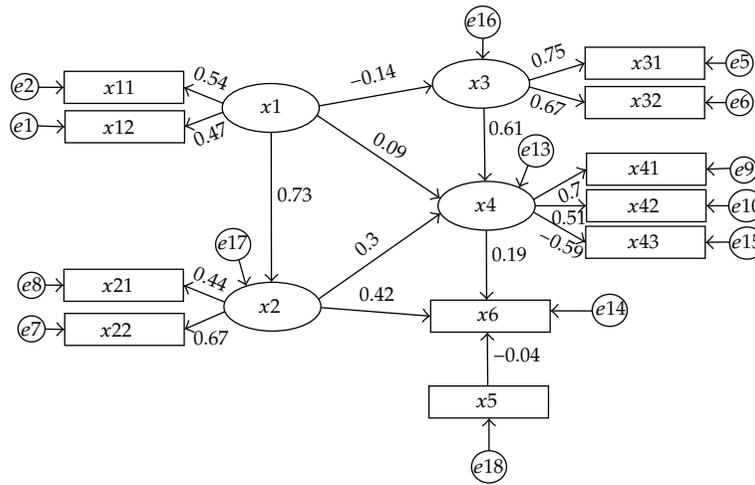


Figure 12: Standard output of model coefficients.

the vector of endogenous observed variables. Λ_y is the coefficient matrix of the endogenous observed variables y on endogenous latent variables η ; ε is the vector of residual error of the endogenous observed variables; x is the vector of exogenous observed variables; Λ_x is the coefficient matrix for the exogenous observed variables x on exogenous latent variables ξ ; and δ is the vector of residual error of the exogenous observed variables.

A whole structural equation modeling has eight coefficient matrixes. The model estimation methods include maximum likelihood (ML), generalized least squares (GLSs), and diagonally weighted least squares (DWLSs). The results of the model estimation are the relation among variables, errors in measure of variables and the portion that cannot be explained by the model, and so forth.

4.2. Selection of Model Variables

Eleven important influencing factors were chosen as the observed variables of structural equation by the correlation analysis. The influencing factors include car ownership, household size, perception of the traffic congestion, parking fee at destination, parking and transferring at park and ride facility, travel time, travel cost, and comfort for riding bus or railway. In order to analyze the effect of past travel behavior on the park and ride intent, the researcher adds another variable, the weekly travel days, for using park and ride. The latent variables are personal information, perception for using park and ride facility, perception of the traffic condition, and perception of the travel.

4.3. Estimation and Analysis of the Model

Amos software is used to establish the structural equation [20]. Figure 11 illustrates the hypothetic relation model for park and ride travel intent by repeatedly estimating parameters and modifying correlations of the models.

Since Amos software is used to denote the error, the residual error symbol in Structure Equation model formulas cannot be expressed by ζ , ε , δ , e . The maximum likelihood

method is used to obtain the goodness-of-fit indices and the standard coefficients for the model (Table 1).

Table 2 shows that the degrees of freedom of the model is 38 and the χ^2 of the model fit index is 26.550. The probability of significance level is 0.919 which refuses the null hypothesis. The statistic indices meet the requirements of relevant standards or critical values except for the normed fit index which is only 0.647. This indicates that the model is fitted to the sample data on the whole.

The standard direct influencing coefficients reflect to what extent the exogenous variables influence the endogenous variables. Their values are denoted by the standardized regression coefficients. Figure 12 shows that the personal information has a significant influence on the perception of the traffic condition with the standard coefficient of 0.73. This indicates that travelers who have fewer cars and household persons will consider that the traffic condition is more congested and parking fee is higher. Personal information has a little effect on the perception of the travel and using park and ride facility. The standard coefficient of the perception of traffic condition on the travel is 0.30. This shows that the poorer the perceptive traffic condition is, the more satisfied the perception for the travel time and cost and comfort for using park and ride is. The standard parameter for the perception of park and ride facility is 0.61. This indicates that the more satisfied the perception of parking and transferring at the facility is, the more satisfied the perceptive travel time and cost for using park and ride is. The standard parameter that the perception of traffic condition influences the travel intent is 0.42. This implies that the more congested the perception of traffic condition is and the higher the perceptive parking fee at destination is, the more inclined the travelers are to choose park and ride and increase the travel times for using park and ride. The standard parameter is 0.19 for the perception of travel of the travel intent which indicates that the more satisfied the travelers are with the perception of travel time, travel cost, and comfort for using park and ride, the more likely they are to choose park and ride and increase the travel times for using park and ride. The standard parameter for the past weekly travel days for using park and ride on the travel intent is -0.04 . This indicates that the past travel behavior has a little effect on travel intent. The previous conclusions show that the poor traffic condition is the main factor that promotes car travelers' usage of park and ride. Improving the service level of park and ride facility and comfort for riding bus or railway will increase the utilization rate of park and ride facilities.

Table 3 shows the indirect influencing relationship among the variables which is caused by the effect of the exogenous observed variables on the other variables via the latent variables or endogenous observed variables. The indirect influencing standard coefficient can be obtained by multiplying the direct influencing standard coefficients. The indirect influencing standard coefficient for the personal information on the park and ride travel intent via the perception of using park and ride facility and travel is -0.016 , and via the perception of travel is 0.017. The indirect influencing standard coefficient for the personal information on the park and ride travel intent via the perception of traffic condition is 0.307, and via the perception of traffic condition and travel is 0.042. Among these indirect relationships, the indirect effect for the personal information via the perception of traffic condition is the biggest. The total indirect influencing effect for the personal information on the travel intent via other variables is 0.35.

The perception of using park and ride facility has an effect of the perception on travel which further affects the park and ride travel intent with the indirect standard coefficient of

Table 1: Variables in structure equation.

Latent variables and variable symbols	Observed variables and variable symbols	Description and coding input values
Personal information X1	Car ownership X11	One car: 1 Two cars: 2 Three cars: 3
	Household size X12	One person: 1 Two persons: 2 Three persons: 3 Four persons: 4 Five persons: 5 Equal to or more than six persons: 6
Perception of traffic condition X2	Traffic congestion X21	Very congested: 1 Relatively congested: 2 Congested: 3 Not congested: 4 Smooth traffic flow: 5
	Parking fee at destination X22	Too high to be tolerated: 1 A little higher: 2 Moderately and acceptable: 3 Cheap: 4
Perception of park and ride facility X3	Perception of parking X31	Strongly satisfied: 1 Relatively satisfied: 2 Satisfied: 3 Unsatisfied: 4 Strongly unsatisfied: 5
	Perception of transferring X32	Strongly satisfied: 1 Relatively satisfied: 2 Satisfied: 3 Unsatisfied: 4 Strongly unsatisfied: 5
Perception of travel X4	Travel time X41	Strongly satisfied: 1 Relatively satisfied: 2 Satisfied: 3 Unsatisfied: 4 Strongly unsatisfied: 5
	Travel cost X42	Strongly satisfied: 1 Relatively satisfied: 2 Satisfied: 3 Unsatisfied: 4 Strongly unsatisfied: 5
	Comfort X43	Too crowded to be tolerated: 1 Relatively crowded: 2 Crowded and acceptable: 3 Comfortable: 4 Very comfortable and enough room: 5
Past weekly travel times for park and ride X5		One day: 1 Two days: 2 Three days: 3 Four days: 4 Five days: 5 Six days: 6 Seven days: 7

Table 1: Continued.

Latent variables and variable symbols	Observed variables and variable symbols	Description and coding input values
Travel intent for park and ride X6		Increase three times: 1 Increase two times: 2 Increase one time: 3 Do not change: 4 Decrease one time: 5 Decrease two times: 6 Decrease three times: 7

Table 2: Goodness-of-fit measures.

Measure	Meaning	Model results	Standards or critical values
χ^2	χ^2	26.550/38 ($P = 0.919$)	$P > 0.05$
CMIN/DF	χ^2 /degrees of freedom	0.699	<3.00
GFI	Goodness-of-fit index	0.949	>0.90
AGFI	Adjusted goodness-of-fit index	0.912	>0.90
NFI	Normed fit index	0.647	>0.90
RMSEA	Root mean square error of approximation	0.000	<0.08
CFI	Comparative fit index	1.000	>0.90
PGFI	Parsimony goodness-of-fit index	0.547	>0.50

0.12. The indirect influencing coefficient of the perception of traffic condition on park and ride travel intent via the perception of travel is 0.06 (Table 4).

These show that various influencing factors have a direct and indirect effect on travel intent based on the interaction among variables. The perception of traffic condition and travel has direct effect on the park and ride travel intent. Personal information has indirect effect on the park and ride travel intent via perception of using park and ride facility, traffic condition, and travel. The perceptions of using park and ride facility and traffic condition have indirect effect on the park and ride travel intent via the perception of travel.

5. Conclusions

Many influencing factors impact the multimodal modes choice for travelers. Park and ride is a multimodal travel behavior. The past researches mainly focused on the relationship between the influencing factors and travel mode choice based on the behavior survey and the disaggregated model. This paper, based on the park and ride survey conducted in Beijing, China, has studied the influence of perception of various factors on the park and ride behavior intent and analyzed the relationship between the perception of the factors and travel intent for park and ride by using the structural equation modeling. The research findings suggest that park and ride is a passive travel behavior which mainly caused by the poor traffic condition. Meanwhile, improving the service level of the park and ride facility and comfort for riding bus or railway will increase the utilization rate for park and ride. All these factors interact with each other. The research also indicates that the perception of various factors has

Table 3: The standard direct relationship.

	Personal information X1	Perception of park and ride facility X3	Perception of traffic condition X2	Perception of travel X4
Perception of park and ride facility X3	-0.14	0.00	0.00	0.00
Perception of traffic condition X2	0.73	0.00	0.00	0.00
Perception of travel X4	0.09	0.61	0.30	0.00
Travel intent X6	0.00	0.00	0.42	0.19

Table 4: The standard indirect relationship.

	Personal information X1	Perception of park and ride facility X3	Perception of traffic condition X2
Perception of transferring X32	-0.09	0.00	0.00
Perception of parking X31	-0.10	0.00	0.00
Comfort X43	-0.13	-0.36	-0.18
Travel intent X6	0.35	0.12	0.06
Travel cost X42	0.12	0.31	0.16
Travel time X41	0.16	0.42	0.21
Traffic congestion X21	0.32	0.00	0.00
Parking fee X22	0.49	0.00	0.00

both direct and indirect effects on travel intent. This research is conducive to the behavior analysis and policy making for park and ride.

Acknowledgments

This research was supported by Beijing Key Laboratory of Traffic Engineering Fund and National Basic Research Program of China (no. 2012CB725403). The authors would like thank H. Guan and Hao Chen for supervising the process. They are very grateful for the comments from the anonymous reviewers.

References

- [1] A. R. Hole, "Forecasting the demand for an employee Park and Ride service using commuters' stated choices," *Transport Policy*, vol. 11, no. 4, pp. 355–362, 2004.
- [2] I. D. M. Bos, R. E. C. M. van der Heijden, E. J. E. Molin, and H. J. P. Timmermans, "The choice of Park and Ride facilities: an analysis using a context-dependent hierarchical choice experiment," *Environment and Planning A*, vol. 36, no. 9, pp. 1673–1686, 2004.
- [3] D. B. Hess, "The effect of free parking on commuter mode choice: evidence from travel diary data," in *Proceedings of the Annual Meeting CD-ROM*, Transportation Research Board, Washington, DC, USA, January 2000.
- [4] H. Ying and H. Xiang, "Study on influence factors and demand willingness of Park and Ride," in *Proceedings of the 2nd International Conference on Intelligent Computing Technology and Automation (ICICTA '09)*, pp. 664–667, October 2009.
- [5] B. H. He, W. N. He, and M. W. He, "The attitude and preference of traveler to the Park & Ride facilities: a case study in Nanjing, China," in *Proceedings of the 8th International Conference on Traffic and Transportation Studies (ICTTS '12)*, vol. 43, pp. 294–301, 2012.

- [6] T. F. Golob, "Joint models of attitudes and behavior in evaluation of the San Diego I-15 congestion pricing project," *Transportation Research Part A*, vol. 35, no. 6, pp. 495–514, 2001.
- [7] R. Sakano and J. Benjamin, "A structural equation analysis of stated travel by commuter rail," in *Proceedings of the Annual Meeting CD-ROM*, Transportation Research Board, Washington, DC, USA, January 2000.
- [8] C. H. Lin, P. J. Sher, and H. Y. Shih, "Past progress and future directions in conceptualizing customer perceived value," *International Journal of Service Industry Management*, vol. 16, no. 4, pp. 318–336, 2005.
- [9] M. M. Rahaman, Md. Abdullah, and A. Rahman, "Measuring service quality using SERVQUAL model: a study on PCBs (private commercial banks) in Bangladesh," *Business Management Dynamics*, vol. 1, no. 1, pp. 1–11, 2011.
- [10] A. Yüksel, "Tourist shopping habitat: effects on emotions, shopping value and behaviours," *Tourism Management*, vol. 28, no. 1, pp. 58–69, 2007.
- [11] D. A. Hensher, P. O. Barnard, and T. P. Truong, "The Role of stated preference methods in studies of travel choice," *Journal of Transport Economics and Policy*, vol. 22, no. 1, pp. 45–58, 1988.
- [12] X. M. Liu, H. Z. Guan, and H. M. Qin, "Modeling park-and-ride choice behavior based on stated-preference survey," in *Proceedings of the Annual Meeting CD-ROM*, Transportation Research Board, Washington, DC, USA, January 2006.
- [13] S. R. Jaeger and J. M. Rose, "Stated choice experimentation, contextual influences and food choice: a case study," *Food Quality and Preference*, vol. 19, no. 6, pp. 539–564, 2008.
- [14] T. F. Golob, "Structural equation modeling for travel behavior research," *Transportation Research Part B*, vol. 37, no. 1, pp. 1–25, 2003.
- [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] W. Wang, H. Guo, H. Bubb, and K. Ikeuchi, "Numerical simulation and analysis procedure for model-based digital driving dependability in intelligent transport system," *KSCE Journal of Civil Engineering*, vol. 15, no. 5, pp. 891–898, 2011.
- [17] W. H. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, Communications Press, Beijing, China, 2012.
- [18] D. Kaplan, *Structural Equation Modeling: Foundations and Extensions*, Sage Publications, 2000.
- [19] R. B. Kline, *Principles and Practice of Structural Equation Modeling*, Guilford Press, 3rd edition, 2010.
- [20] B. M. Byrne, *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 2001.

Research Article

A Numerical Model for Railroad Freight Car-to-Car End Impact

Chao Chen, Mei Han, and Yanhui Han

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Chao Chen, chen.q@163.com

Received 11 September 2012; Revised 9 November 2012; Accepted 21 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Chao 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.

A numerical model based on Lagrange-D'Alembert principle is proposed for car-to-car end impact in this paper. In the numerical model, the friction forces are treated by using local linearization model when solving the differential equations. A computer program has been developed for the numerical model based on Runge-Kutta fourth-order method. The results are compared with the Multibody Dynamics/Kinematics software SIMPACK results and they are close. The ladings' relative displacement to struck car and the relative displacement between two ladings get larger as impact speed increases. There is no displacement between two ladings when the contact surfaces have the same friction coefficient.

1. Introduction

The freight damage incurred during railroad transportation is a serious economic and safety problem. The railroad freight car's dynamic characteristic leads to most of the freight damage. The dynamics of railroad car and freight damage can be divided into two groups:

- (1) during the marshalling operation in train yard, the car-to-car end impacts from coupling cause high car and lading acceleration;
- (2) the car-body vibrations come from track irregularities and some extra forces, such as the wind, and so forth.

Most of the damage is attributed to car-to-car end impacts in the marshalling yard, so more focus is given on it when working out the load support and load securement method. Railroad freight car impact tests are always carried out for checking if the method can ensure transportation safety and no damage to the ladings. A railroad freight car impact test usually needs a lot of work to do; it needs much workforce, material resources, and financial support. Most of the time, carrying out an impact test will lead to disorder and break-off

transportation. Compared with impact test, numerical simulation is a more economical and faster method of investigating the effect on loadings when coupled.

Car-to-car end impact is a special multibody dynamic problem between railroad freight cars. Investigation into the multibody dynamic has been carried out in the works [1–9]. Later, mathematical models are derived in [10] for studying the effect of impact on packaging. At the same time, numerical methods need to be developed for solving mathematical models. Euler tangent method, Newmark- β method, Wilson- θ method, and Runge-Kutta fourth-order method are developed and widely applied in solving mathematical models [11–17]. Runge-Kutta fourth-order method means that the truncation error per step is $O(h^5)$. It is an important numerical method used extensively in engineering problems for solving first-order differential equations.

Draft gear is the most important component of a freight car during impact. Its performance is investigated by mechanics dynamics software in [18–21]. The draft gear's characteristic is analyzed under different impact speeds. The force versus draft gear travel of the Chinese MT-2 under impact speed of 5~8 km/h is given by simulation and test.

In the paper, the second-order differential equations of the car-to-car end impact are converted to first-order differential equations and solved by using Runge-Kutta 4th order method.

2. Draft Gear Interaction Process

Railroad car-to-car end impacts usually occur in train yard, and most of the time the struck car is static when coupled with the striking car. The draft gear is an important component for reducing freight and car damage during car-to-car end impacts.

MT-2 friction draft gear is widely used in the class 70 t universal freight cars in China. This draft gear is composed by springs and friction mechanism; when it is compressed, part of kinetic energy is converted to friction energy and part of kinetic energy is converted to potential energy. MT-2 draft gear has different force versus travel characteristic curves when loading and unloading. In Figure 1, the irreversible force versus draft gear travel characteristic curve is shown [22].

As shown in Figure 1, Δx is draft gear travel and Δv is speed difference between striking car and struck car. $\Delta v > 0$ means draft gear loading process and $\Delta v < 0$ means unloading process. Figure 1 shows that the resistant force in loading process is larger than unloading process. In the numerical calculation program, draft gear force versus travel characteristic curve is based on the test results, and the force is calculated by linear interpolation. MT-2 draft gear force versus travel characteristic curves under impact speeds of 5 km/h, 6 km/h, 7 km/h, and 8 km/h are presented in the appendix.

3. Car-to-Car End Impact Dynamic Models

3.1. Railroad Freight Car Impact System

Railroad freight car impact test is using a striking car with a certain speed running to a static struck car and collides. The longitudinal status of the loadings and struck car is mainly observed during impact for checking the loading support and loading secure method. The method must ensure transportation safety and no loading damage.

The assumptions in models are

- (1) the wind acting on the striking car and struck car, and the rolling resistance between wheel and rail are neglected;

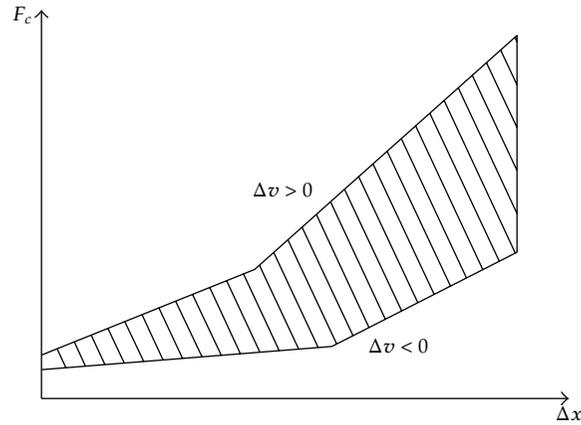


Figure 1: Draft gear's force and travel characteristic.

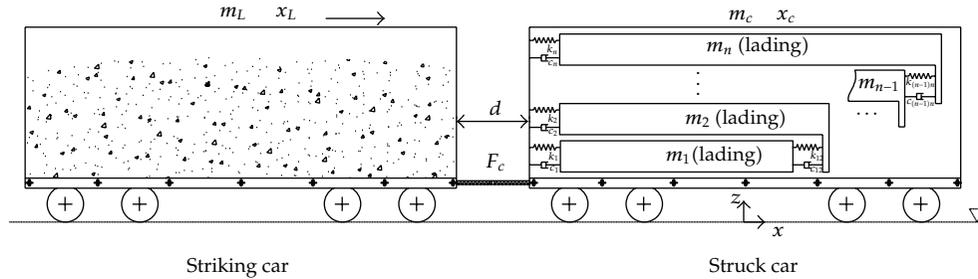


Figure 2: Car-to-car end impact dynamic system.

- (2) the car-body deformations during impact are neglected;
- (3) the car-body vertical bounce, yaw, pitch, and sway vibrations are neglected;
- (4) no lateral forces between ladings.

Sometimes more than one lading are loaded on freight car. There are longitudinal forces between ladings and car, between adjacent ladings. Figure 2 shows the railroad freight car-to-car end impact dynamic system, where m , x represent mass and displacement, L represents striking car, c represents struck car, $1, 2, \dots, n$ represent ladings, k_n, c_n are the stiffness and damping coefficients between ladings and car, $k_{(n-1)n}, c_{(n-1)n}$ are the stiffness and damping coefficients between ladings n and $(n - 1)$, F_c is draft gear force, and d is distance between two cars.

3.2. Longitudinal Dynamic Equations

The striking car, struck car and ladings in the impact dynamic system are treated as mass elements. According to the assumptions in Section 3.1, the external forces, constraint forces and inertial forces are an equivalent static system based on Lagrange-D'Alembert principle. Then, universal longitudinal dynamic equations can be derived for each mass element.

External force acting on the striking car is the draft gear force. So the differential equation for striking car is

$$m_L \ddot{x}_L + F_c = 0, \quad (3.1)$$

where m_L is gross weight of the striking car and F_c is calculated by the following equation:

$$F_c = \begin{cases} F_1(\Delta x) & \Delta v > 0, \\ F_2(\Delta x) & \Delta v < 0. \end{cases} \quad (3.2)$$

The acceleration and velocity initial values of the striking car are $\ddot{x}_L = 0$, $\dot{x}_L = v_L$; the initial value of the draft gear force is $F_c = 0$.

The external forces acting on the struck car are the draft gear force and forces between ladings and car,

$$m_c \ddot{x}_c + \sum_{i=1}^n c_i (\dot{x}_c - \dot{x}_i) + \sum_{i=1}^n k_i (x_c - x_i) - F_c = 0, \quad (3.3)$$

where m_c is the struck car tare weight and the acceleration and velocity initial values of struck car are $\ddot{x}_c = 0$, $\dot{x}_c = 0$.

External forces acting on lading i are the forces between lading i and car, between adjacent ladings,

$$\begin{aligned} & m_i \ddot{x}_i - c_i (\dot{x}_c - \dot{x}_i) - k_i (x_c - x_i) - c_{(i-1)i} (\dot{x}_{i-1} - \dot{x}_i) - k_{(i-1)i} (x_{i-1} - x_i) \\ & \quad + c_{i(i+1)} (\dot{x}_i - \dot{x}_{i+1}) + k_{i(i+1)} (x_i - x_{i+1}) = 0, \quad i \in [2, n-1], \\ & m_1 \ddot{x}_1 - c_1 (\dot{x}_c - \dot{x}_1) - k_1 (x_c - x_1) + c_{12} (\dot{x}_1 - \dot{x}_2) \\ & \quad + k_{12} (x_1 - x_2) = 0, \quad i = 1, \\ & m_n \ddot{x}_n - c_n (\dot{x}_c - \dot{x}_n) - k_n (x_c - x_n) - c_{(n-1)n} (\dot{x}_{n-1} - \dot{x}_n) \\ & \quad - k_{(n-1)n} (x_{n-1} - x_n) = 0, \quad i = n, \end{aligned} \quad (3.4)$$

where the acceleration and velocity initial values of lading i are $\ddot{x}_i = 0$, $\dot{x}_i = 0$.

4. Double-Stack Loading Impact Models

Double-stack loading method in gondola car is proposed as an example for analyzing longitudinal relation between ladings and car. It shows that the numerical method applied in railroad freight car-to-car end impact simulation. In the model, the striking car and struck car are the same type and have the same draft gears.

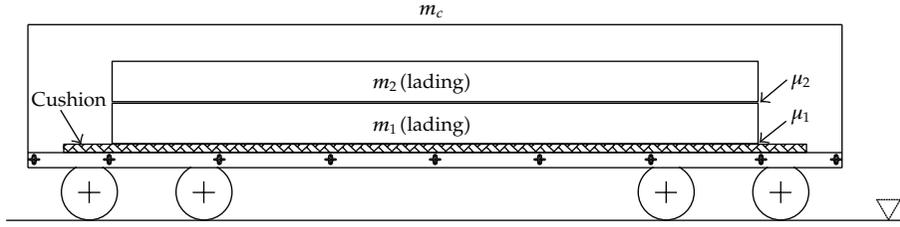


Figure 3: Load support and load securement method.

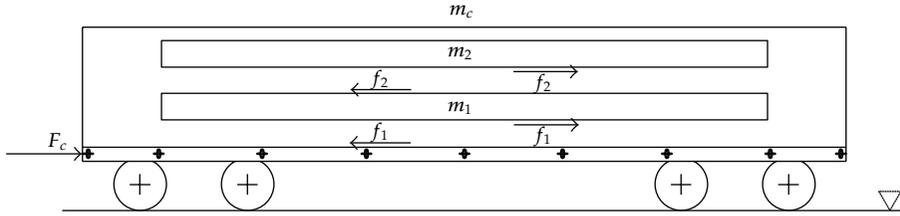


Figure 4: Force analysis.

4.1. Load Support and Load Securement Method and Force Analysis

Figure 3 shows the double-stack loading in a 70 t class gondola car. The securement method is using friction cushion to enlarge friction force.

The above struck car system includes three mass elements that are the struck car, 1st lading and 2nd lading. Forces acting on struck car are draft gear force and friction force from 1st lading; forces acting on the 1st lading are friction forces from struck car and the 2nd lading; force acting on the 2nd lading is friction force from 1st lading. The force analysis is illustrated in Figure 4.

4.2. Dynamic Equations of Motion

The universal longitudinal dynamic equations in Section 3.2 can be rewritten based on the force analysis

$$\begin{aligned}
 \text{striking car: } m_L \ddot{x}_L &= -F_c, \\
 \text{struck car: } m_c \ddot{x}_c &= F_c - f_1, \\
 \text{1st lading: } m_1 \ddot{x}_1 &= f_1 - f_2, \\
 \text{2nd lading: } m_2 \ddot{x}_2 &= f_2.
 \end{aligned} \tag{4.1}$$

As striking car and struck car have the same draft gear type, so the travel of one draft gear is given as

$$\Delta x = \frac{(x_L - x_c)}{2}. \tag{4.2}$$

The speed difference between striking car and struck car is given as

$$\Delta v = \dot{x}_L - \dot{x}_c. \quad (4.3)$$

The model has two one-dimensional friction elements: one is between 1st lading and car-body and the other is between 1st lading and 2nd lading. The one-dimensional friction element's friction force direction is dependent on the direction of the relative sliding velocity. Figure 5(a) shows that the direction of friction force changes abruptly as the direction of relative velocity changes. More calculation time is needed near the zero-point, and even the differential equations cannot be integrated at zero-point. The friction element is treated by using a local linearization model [23], which uses a parameter v_0 called switching speed. Figure 5(b) illustrates linear relationship between friction force and relative velocity, and the friction force is given as

$$f = \begin{cases} \mu mg & \Delta v > v_0, \\ \frac{\Delta v}{v_0} \mu mg & |\Delta v| \leq v_0, \\ -\mu mg & \Delta v < -v_0. \end{cases} \quad (4.4)$$

4.3. Solution Methodology

For using Runge-Kutta fourth-order method, second-order differential equations need to be rewritten in the form of first-order differential equations. In general, to solve first-order differential equation $\dot{y} = f(x, y)$, Runge-Kutta fourth-order method is given as

$$\begin{aligned} y_{i+1} &= y_i + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4), \\ K_1 &= f(x_i, y_i), \\ K_2 &= f\left(x_i + \frac{h}{2}, y_i + \frac{h}{2}K_1\right), \\ K_3 &= f\left(x_i + \frac{h}{2}, y_i + \frac{h}{2}K_2\right), \\ K_4 &= f(x_i + h, y_i + hK_3), \end{aligned} \quad (4.5)$$

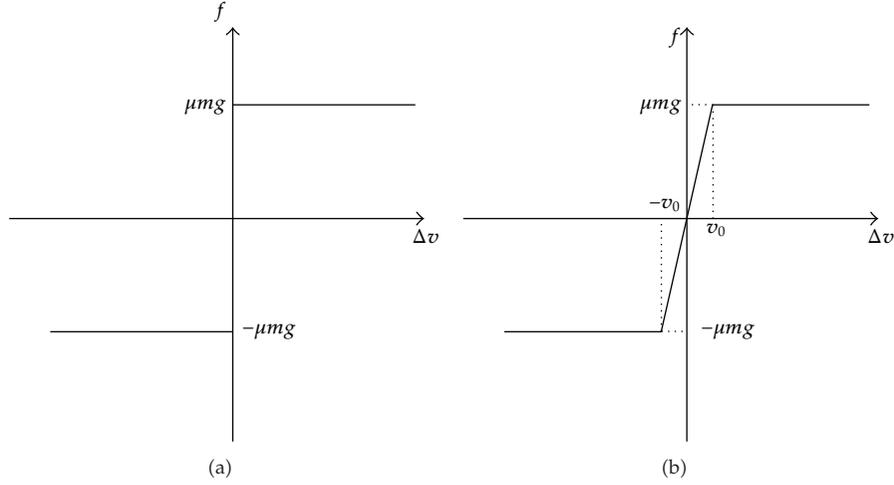
where h is the step size.

Let $x_{dL} = \dot{x}_L$ in the striking car's second-order differential equation, then the reduced-order differential equations are given as

$$\begin{aligned} \dot{x}_{dL} &= -\frac{F_c}{m_L}, \\ \dot{x}_L &= x_{dL}. \end{aligned} \quad (4.6)$$

Table 1: Numerical model parameters.

Case	m_L/t	m_c/t	m_1/t	m_2/t	μ_1	μ_2	h/s
1	92.5	22.5	30	30	0.4	0.3	0.00001
2	92.5	22.5	30	30	0.4	0.4	0.00001
3	92.5	22.5	30	20	0.4	0.3	0.00001

**Figure 5:** One-dimensional friction element model.

In the same way, the second-order differential equations of other mass elements are given as,

$$\begin{aligned}
 \text{Struck car} & \begin{cases} \dot{x}_{dc} = \frac{F_c - f_1}{m_c}, \\ \dot{x}_c = x_{dc}, \end{cases} \\
 \text{1st lading} & \begin{cases} \dot{x}_{d1} = \frac{f_1 - f_2}{m_1}, \\ \dot{x}_1 = x_{d1}, \end{cases} \\
 \text{2nd lading} & \begin{cases} \dot{x}_{d2} = \frac{f_2}{m_2}, \\ \dot{x}_2 = x_{d2}. \end{cases}
 \end{aligned} \tag{4.7}$$

4.4. Numerical Results and Discussion

The type of draft gear is MT-2 in the dynamic models; the impact speeds are 5 km/h, 6 km/h, 7 km/h, and 8 km/h. The simulation is for studying relative displacement of 1st lading and 2nd lading, which are secured by the friction cushion. Case 1 only has friction cushion between 1st lading and car floor; Case 2 has friction cushion between 1st lading and car floor, and between two ladings; the friction cushion in Case 3 is the same as Case 1, but two ladings have different weight. The parameters in numerical model are given in Table 1.

A virtual model is built in Multibody Dynamics/Kinematics software SIMPACK which has the same parameters as Case 1 to validate the model and numerical method. It

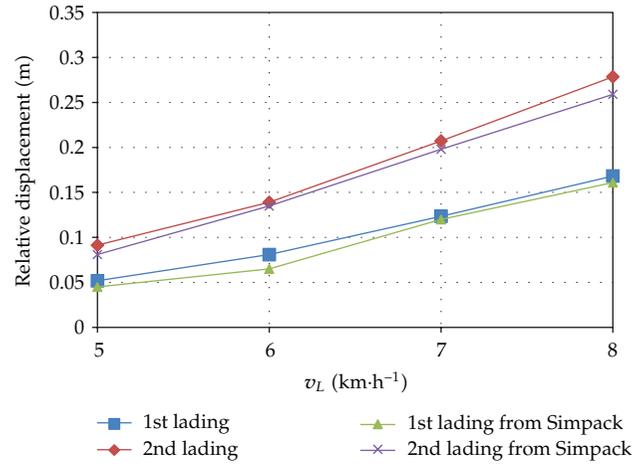


Figure 6: Results under $m_1 = 30$ t, $m_2 = 30$ t, $\mu_1 = 0.4$, and $\mu_2 = 0.3$.

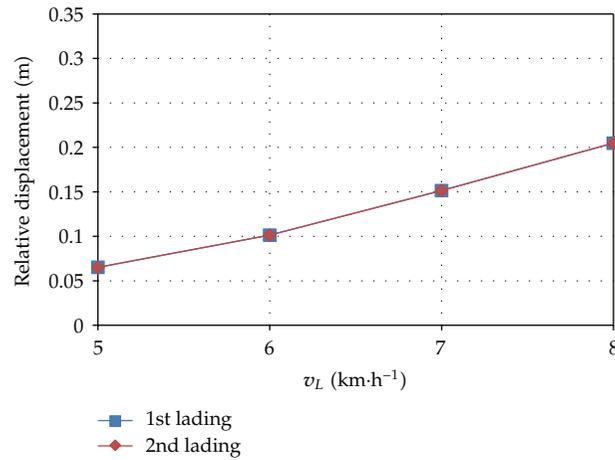


Figure 7: Results under $m_1 = 30$ t, $m_2 = 30$ t, $\mu_1 = 0.4$, and $\mu_2 = 0.4$.

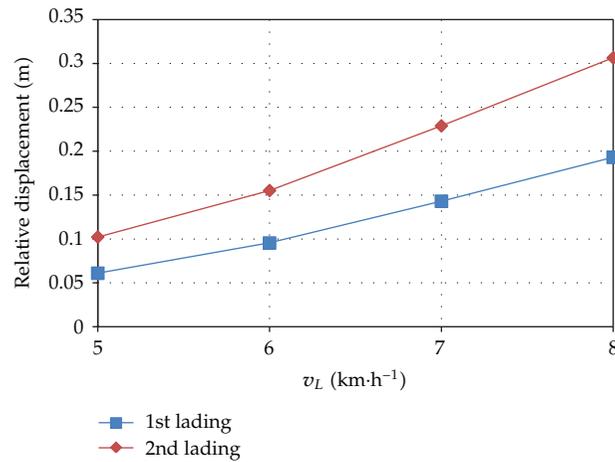
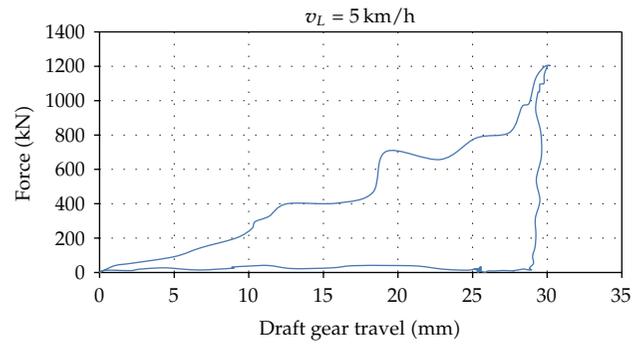
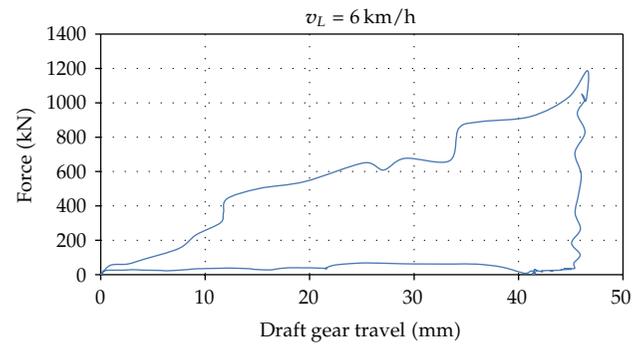


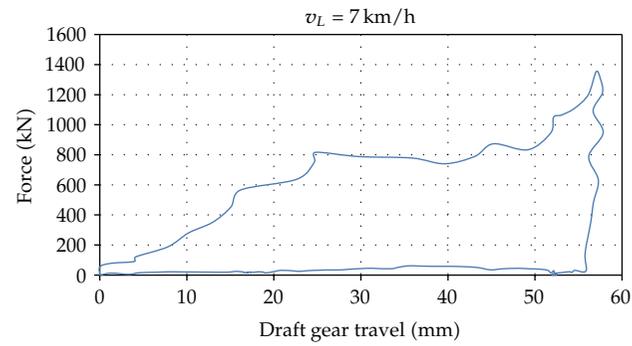
Figure 8: Results under $m_1 = 30$ t, $m_2 = 20$ t, $\mu_1 = 0.4$, and $\mu_2 = 0.3$.



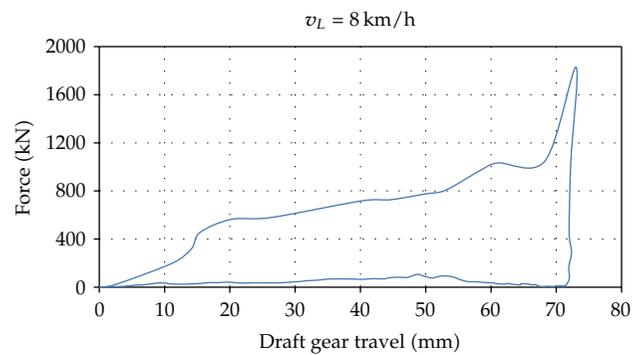
(a)



(b)



(c)



(d)

Figure 9: Draft gear force and travel characteristic curves. v_L is the velocity of striking car just before impact.

can be observed from Figure 6 that there is an excellent agreement between the results from Runge-Kutta fourth-order method and SIMPACK. The displacement between ladings and car-body and between 1st lading and 2nd lading increased with impact speeds.

Figure 7 shows the displacement between ladings and car-body under Case 2 versus different impact speeds. The displacement between ladings and car-body increased with impact speeds, but the displacement between 1st lading and 2nd lading is zero as two surfaces have the same friction coefficient.

In Case 3, the weight of 2nd lading is less than that in Case 1, so the gross weight of struck car reduced. The displacement between ladings and car-body versus impact speeds are illustrated in Figure 8. The displacements between ladings and car-body are increased compared with Case 1 under the same impact speed.

5. Conclusion

In this paper, railroad freight car-to-car end impact system and the influence on ladings are considered. To derive differential equations of the system motion, forces acting in the system are analyzed and Lagrange-D'Alembert principle is used. The obtained solution of the differential equations by Runge-Kutta fourth-order method is close to the results from SIMPACK. Based on numerical results from double-stack model, it is concluded that the higher the impact speed, the larger the ladings' relative displacement to struck car. The lower weight the ladings, the larger the ladings' relative displacement to struck car. There is no displacement between two ladings if they have the same friction coefficient between struck car and 1st lading and between 1st lading and 2nd lading.

Appendix

See Figures 9(a)–9(d).

Acknowledgment

This paper is supported by “the Fundamental Research Funds for the Central Universities” (2011JBM246).

References

- [1] W. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, China Communications Press, Beijing, China, 2012.
- [2] H. J. Fletcher, L. Rongved, and E. Y. Yu, “Dynamics analysis of a two-body gravitationally oriented satellite,” *Bell System Technical Journal*, vol. 42, no. 5, pp. 2239–2266, 1963.
- [3] W. W. Hooker and G. Margulies, “The dynamical attitude equations for an n -body satellite,” *American Astronautical Society. Journal of the Astronautical Sciences*, vol. 12, pp. 123–128, 1965.
- [4] W. W. Hooker, “A set of r-dynamical attitude equations for an arbitrary n -body satellite having rotational degrees of freedom,” *American Institute of Aeronautics and Astronautics Journal*, vol. 8, no. 7, pp. 1205–1207, 1970.
- [5] R. Schwertassek and R. E. Roberson, “A state-space dynamical representation for multibody mechanical systems part I: systems with tree configuration,” *Acta Mechanica*, vol. 50, no. 3-4, pp. 141–161, 1984.
- [6] R. E. Roberson and R. Schwertassek, *Dynamics of Multibody Systems*, Springer, Berlin, Germany, 1988.
- [7] 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.

- [8] R. E. Roberson and W. Wittenburg, "A dynamical formalism for an arbitrary number of interconnected rigid bodies with reference to the problem of satellite attitude control," in *Proceedings of the 3rd International Federation of Automatic Control Congress (IFAC '66)*, London, UK, 1966.
- [9] J. Wittenburg, *Dynamics of Systems of Rigid Bodies*, Teubner, Stuttgart, Germany, 1977.
- [10] R. V. Dukkipati, *Vehicle Dynamics*, CRC Press, Boca Raton, Fla, USA, 2000.
- [11] N. M. Newmark, "A method of computation for structural dynamics," *Journal of the Engineering Mechanical Division*, vol. 85, no. 2, pp. 67–94, 1959.
- [12] E. L. Wilson, I. Farhoomand, and K. J. Bathe, "Nonlinear dynamic analysis of complex structure," *Earthquake Engineering and Structural Dynamics*, vol. 1, no. 3, pp. 241–252, 1973.
- [13] K. J. Bath and E. L. Wilson, *Numerical Methods in Finite Element Analysis*, Prentice Hall, New York, NY, USA, 1976.
- [14] E. Süli and D. F. Mayers, *An Introduction to Numerical Analysis*, Cambridge University Press, Cambridge, UK, 2003.
- [15] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer, New York, NY, USA, 3rd edition, 2002.
- [16] G. Da-Hu and S. A. Meguid, "Stability of Runge-Kutta methods for delay differential systems with multiple delays," *IMA Journal of Numerical Analysis*, vol. 19, no. 3, pp. 349–356, 1999.
- [17] W. Zhiqiao, *L-Stable Methods for Numerical Solution of Structural Dynamics Equations and Multibody Dynamics Equations*, National University of Defense Technology, 2009.
- [18] T. Guangrong, *Study on System Dynamics of Heavy Haul Train*, Southwest Jiaotong University, 2009.
- [19] W. Chengguo, M. Dawei, W. Xuejun, and L. Lan, "Research on structures and performances of heavy haul buffer by numerical simulation," *Railway Locomotive & Car*, vol. 29, no. 5, pp. 1–4, 2009.
- [20] H. Guoliang, L. Fengtao, W. Xuejun, and W. Chengguo, "Characteristics research in spring friction draft gear of heavy haul freight cars," *Lubrication Engineering*, vol. 34, no. 7, pp. 69–73, 2009.
- [21] L. Ming, H. Guoliang, W. Xuejun, and W. Chengguo, "Simulation analysis and experiment research into two typical draft gear of heavy haul freight cars," *Coal Mine Machinery*, vol. 30, no. 9, pp. 76–78, 2009.
- [22] H. Yunhua, L. Fu, and F. Maohai, "Research on the characteristics of vehicle buffers," *China Railway Science*, vol. 26, no. 1, pp. 95–99, 2005.
- [23] D. Karnopp, "Computer simulation of stick-slip friction in mechanical dynamic systems," *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, vol. 107, no. 1, pp. 101–103, 1985.

Research Article

Traffic Behavior in CA Model of Vehicular Traffic through a Series of Signals

Kazuhiro Tobita and Takashi Nagatani

Department of Mechanical Engineering, Shizuoka University, Hamamatsu 432-8561, Japan

Correspondence should be addressed to Takashi Nagatani, tmtnaga@ipc.shizuoka.ac.jp

Received 11 September 2012; Accepted 14 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 K. Tobita and T. Nagatani. 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 study the traffic states and fundamental diagram of vehicular traffic controlled by a series of traffic lights using a deterministic cellular automaton (CA) model. The CA model is not described by a set of rules but is given by a difference equation. The vehicular traffic varies highly with both signal's characteristics and vehicular density. The dependence of fundamental diagram on the signal's characteristics is derived. At a low value of cycle time, the fundamental diagram displays the typical trapezoid, while it shows a triangle at a high value of cycle time. The dynamic transitions among distinct traffic states depend greatly on the signal's characteristics. The dependence of the transition points on the cycle time split and offset time is found.

1. Introduction

Mobility is nowadays one of the most significant ingredients of a modern society. Recently, transportation problems have attracted much attention in the fields of physics [1–4]. Physics and other sciences meet at the frontier area of interdisciplinary research. The traffic flow, pedestrian flow, and bus-route problem have been studied from a point of view of statistical mechanics and nonlinear dynamics [5–7]. The interesting dynamic behaviors have been found in the transportation system. The jams, chaos, and pattern formation are typical signatures of the complex behavior of transportation [8, 9]. The cellular automaton (CA) model has been used extensively for the traffic dynamics [1, 3]. The CA model for traffic flow is a typical system of discrete dynamics.

The traffic light is an essential element for managing the transportation network. In urban traffic, vehicles are controlled by traffic lights to give priority for a road because the city traffic networks often exceed the capacity and one avoids a collision between vehicles. Brockfeld et al. have studied optimizing traffic lights for city traffic by using a CA traffic

model [10]. They have clarified the effect of signal control strategy on vehicular traffic. Also, they have shown that the city traffic controlled by traffic lights can be reduced to a simpler problem of a single-lane roadway. D. W. Huang and W. N. Huang have studied the traffic flow controlled by signals on a single-lane roadway by using Nagel-Schreckenberg model [11]. Sasaki and Nagatani have investigated the traffic flow on a single-lane roadway with traffic lights by using the optimal velocity model [12]. They have derived the relationship between the road capacity and jamming transition. Until now, one has studied the periodic traffic controlled by a few traffic lights. It has been concluded that the periodic traffic does not depend on the number of traffic lights. Very recently, Lammer and Helbing have studied the vehicular flow by the self-control of traffic signals in urban road networks [13].

In real traffic, the vehicular traffic depends highly on the configuration of traffic lights and the priority of roadways. In the dilute limit of vehicular density, a few works have been done for the traffic of vehicles moving through an infinite series of traffic lights with the same interval. The effect of cycle time on vehicular traffic has been clarified by using the nonlinear-map models [14–18]. Also, it has been shown that the heterogeneity of signal's interval and irregular split have the important effects on vehicular traffic [19].

The deterministic CA model has been presented for the vehicular traffic through a series of traffic lights at the synchronized strategy for various values of vehicular density [20]. The dependence of the travel time on both cycle time and density has been clarified. The effect of a slow vehicle on the traffic flow through a series of signals has been investigated [21]. Also, it has been shown that the fundamental diagram depends a little on the cycle time for an intermediate value of the cycle time [20]. It has been found that the dynamic transition occurs from the undersaturated traffic, through the saturated traffic, to the oversaturated traffic [20]. However, it is little known how the dynamic transition depends on the signal's characteristics for the traffic flow through a series of signals. Also, it is unknown how the dynamic transition varies with the cycle time at the green-wave strategy. It will be necessary and important to study the dynamic transition and the fundamental diagram for the traffic flow through a series of signals at the green-wave strategy.

Until now, the traffic flow controlled by a few signals has been studied. Also, the traffic flow through the series of signals controlled by split 0.5 and zero offset has been investigated with the use of CA model. The effects of split and offset on traffic flow through the series of signals are known little. Also, the dependence of the fundamental diagram on the cycle time, the split, and the offset is unclear.

In order to study the effect of offset on the traffic flow, it is necessary to take into account the series of signals because the phase shift varies from signal to signal at the green-wave strategy.

In this paper, we take into account the split and offset. We clarify the effect of split and offset on the traffic flow. We study the fundamental diagram and dynamic transitions for traffic flow through a series of traffic lights at the synchronized and green-wave strategies by using the CA model described by the difference equation. We clarify the dependence of dynamic transition on the signal's characteristics. We show how the fundamental diagram and transition points vary with the cycle time, split, and phase shift of signals. Especially, we investigate the dynamic transitions and fundamental diagram at a high value of the cycle time.

2. CA Model and Difference Equation

We consider the flow of vehicles going through the series of traffic lights on one-dimensional lattice. Each vehicle does not pass over other vehicles. The traffic lights are positioned homogeneously on a roadway. The interval between signals has a constant value and is given by l . All traffic lights change from red (green) to green (red) with a fixed time period $(1 - s_p)t_s$ ($s_p t_s$). The period of green is $s_p t_s$ and the period of red is $(1 - s_p)t_s$. Time t_s is called the cycle time and fraction s_p represents the split which indicates the ratio of green time to cycle time.

We apply the Fukui-Ishibashi model to the vehicular motion [22]. The Fukui-Ishibashi model is not stochastic but deterministic. We use the deterministic CA model in place of the stochastic Nagel-Schreckenberg model [3] because the dynamic transition occurs clearly in the deterministic model. We extend the deterministic CA model to take into account traffic lights. We define the position of vehicle i at time t as $x_i(t)$ where x , i , and t are an integer. The velocity takes the integer value ranging from 0 to v_{\max} . The velocity depends on the headway. If headway $\Delta x_i(t) (= x_{i+1}(t) - x_i(t))$ is larger than the maximum velocity, the vehicle moves with the maximum velocity. If the headway is less than the maximum velocity, the vehicle moves with velocity $\Delta x_i(t) - 1$.

When a vehicle arrives at a traffic light and the traffic light is red, the vehicle stops at the position of the traffic light. Then, when the traffic light changes from red to green, the vehicle goes ahead. On the other hand, when a vehicle arrives at a traffic light and the traffic light is green, the vehicle does not stop and goes ahead without changing speed. The position of the closest signal before vehicle i at time t is given by

$$x_{i,s}(t) = \left\{ \text{int} \left(\frac{x_i(t)}{l} \right) + 1 \right\} l. \quad (2.1)$$

We set split as $s_p = 0.5$. Then, the position of vehicle i at time $t + 1$ is given by

$$\begin{aligned} x_i(t + 1) = & \min[x_i(t) + v_{\max}, x_{i+1}(t) - 1] \left\{ 1 - \vartheta \left(\sin \left(\frac{2\pi t}{t_s} \right) \right) \right\} \\ & + \min[x_i(t) + v_{\max}, x_{i+1}(t) - 1, x_{i,s}(t) - 1] \vartheta \left(\sin \left(\frac{2\pi t}{t_s} \right) \right). \end{aligned} \quad (2.2)$$

Here, $\vartheta(t)$ is the step function. It takes one if $t > 0$ and zero if $t \leq 0$. $\min[A, B]$ is a minimum function and takes the minimum value within A and B .

If the signal just before vehicle i is green, $\vartheta(\sin(2\pi t/t_s)) = 0$ and $x_i(t + 1) = \min[x_i(t) + v_{\max}, x_{i+1}(t) - 1]$. Otherwise, if the signal just before vehicle i is red, $\vartheta(\sin(2\pi t/t_s)) = 1$ and $x_i(t + 1) = \min[x_i(t) + v_{\max}, x_{i+1}(t) - 1, x_{i,s}(t) - 1]$. Then, if the headway is larger than v_{\max} , vehicle i stops at site $x_{i,s}(t) - 1$ just before the signal. Also, if $x_{i+1}(t)$ is higher than $x_{i,s}(t)$, vehicle i stops at site $x_{i,s}(t) - 1$ just before the signal. Thus, (2.2) presents the CA model for the vehicular traffic through a series of traffic lights. Equation (2.2) is a single difference equation. Until now, the CA model for the signal traffic has been described by a set of CA rules. However, model (2.2) is of great advantage to simulate the traffic flow because the difference equation is simple.

We extend (2.2) to take into account the phase shift and split. The signal timing is controlled by offset time t_{offset} . The offset time means the difference of phase shifts between

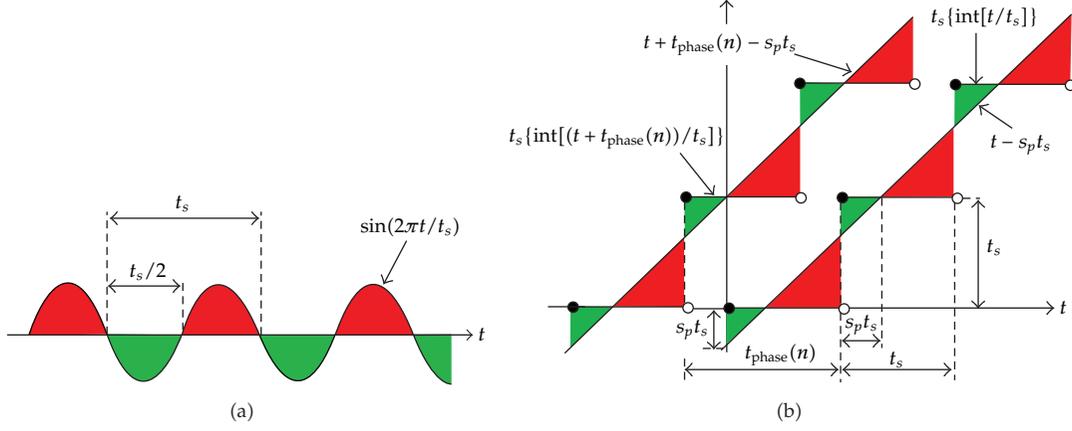


Figure 1: Schematic illustration of periodic functions. (a) Sine wave. (b) Generalized periodic function.

two successive signals. In the green wave (delayed) strategy, the phase shift of signal n is given by $t_{\text{phase}}(n) = n t_{\text{offset}}$ where the phase shift is indicated by $t_{\text{phase}}(n)$. Then, the signal switches from red to green in green wave way. The phase shift increases with signal n .

The position of vehicle i at time $t + 1$ is given by

$$\begin{aligned}
 x_i(t+1) &= \min[x_i(t) + v_{\max}, x_{i+1}(t) - 1] \\
 &\times \left\{ 1 - \vartheta \left[t + t_{\text{phase}}(n) - t_s \left\{ \text{int} \left(\frac{t + t_{\text{phase}}(n)}{t_s} \right) \right\} - s_p t_s \right] \right\} \\
 &+ \min[x_i(t) + v_{\max}, x_{i+1}(t) - 1, x_{i,s}(t) - 1] \\
 &\times \vartheta \left[t + t_{\text{phase}}(n) - t_s \left\{ \text{int} \left(\frac{t + t_{\text{phase}}(n)}{t_s} \right) \right\} - s_p t_s \right].
 \end{aligned} \tag{2.3}$$

Periodic function $\sin(2\pi t/t_s)$ in (2.2) is replaced by generalized periodic function $[t + t_{\text{phase}}(n) - t_s \{ \text{int}((t + t_{\text{phase}}(n))/t_s) \} - s_p t_s]$ in (2.3). In the generalized periodic function, the phase shifts are included explicitly. However, if one extends periodic function $\sin(2\pi t/t_s)$ to take into account the split, one cannot include the split explicitly. Figure 1(a) shows the schematic illustration of traffic sign given by a sine wave in (2.2). The cycle time is given by t_s . The split is $s_p = 0.5$. Figure 1(b) shows the schematic illustration of the generalized periodic function in (2.3). The split is set explicitly by any value s_p . Also, the phase shift is displayed by t_{phase} . The green and red signs are illustrated by the green and red colors, respectively.

When $s_p = 0.5$ and $t_{\text{phase}} = 0$, (2.3) reduces to (2.2). If there are no signals on the roadway, (2.3) reduces to the CA model of Fukui and Ishibashi

$$x_i(t+1) = \min[x_i(t) + v_{\max}, x_{i+1}(t) - 1], \tag{2.4}$$

where v_{\max} is the maximum velocity and an integer.

It will be expected that the vehicular traffic exhibits a complex behavior by dynamic transitions. We study how the vehicular traffic changes by varying the cycle time, split, phase shift, and vehicular density.

Until now, CA models with no signals and with a few signals have been proposed. In the previous work [20], we have presented the CA model with the series of signals at the synchronized strategy. The split and the offset (phase shift) have been not taken into account. Generally, it is necessary and important to take into account the split and the offset (phase shift). In this study, we have proposed the CA model with the series of signals to take into account the split and the offset of signals. The conventional CA model is described in terms of the set of rules. However, our CA model is given by the difference equation (2.3). The difference equation (2.3) is the first for the traffic flow controlled by the series of signals.

3. Simulation Result

We investigate the fundamental diagram, the traffic states, and the dynamic transitions by using (2.3). First, we study the traffic flow at the synchronized strategy for the signal control. Second, we investigate the traffic flow controlled by signals at the green-wave strategy.

We consider the city traffic network controlled by traffic signals. There always exist signals at a crossing. Vehicles move ahead through a series of signals at a street. We simulate the traffic flow on a single-lane roadway with many signals. We set the direction of vehicle flow as x -axis. A signal changes alternately from green (red) to red (green). We ignore the yellow signal. We consider the periodic boundary for the vehicular flow on a single-lane roadway. When vehicles reach the right edge of the roadway, they return the left edge of the roadway. We calculate the positions of all vehicles using (2.3).

3.1. Synchronized Strategy

In the synchronized strategy, all signals change simultaneously. We calculate the current (flow) versus density for various values of cycle time t_s . We set the interval between signals as $l = 40$ and the maximal velocity as $v_{\max} = 4$. We define the dimensionless cycle time as $T_s = t_s v_{\max} / l$. It is the ratio of cycle time to the moving time between a signal and its next signal. This means that the unit of time is the moving time over the interval between two signals. It has an advantage that the dimensionless cycle time does not vary even if the interval between two signals or maximum velocity changes. Also, the dimensionless offset time is defined as $\tau = t_{\text{offset}} v_{\max} / l$. Figure 2(a) shows the plots of mean current against density for dimensionless cycle time $T_s = 3.0$ where the split is $s_p = 0.5$, the offset time is $\tau = 0.0$, and the road length is $L = 4000$. The solid curve indicates the current profile. For comparison, the current for the vehicular traffic with no signals is shown by the chain line. The current profile changes from the triangle to the trapezoid by adding the signals. The traffic states change from the undersaturated traffic, through the saturated traffic, to the oversaturated traffic. In the unsaturated traffic region between points 0 – b , the current increases with density, two distinct states occur, and the traffic state changes at point a . In the saturated traffic region between points b – c , the current keeps a constant value. In the oversaturated traffic region between points c – 1 , the current decreases with increasing density, two distinct states appear, and the traffic state changes at point d . Thus, there exist five distinct traffic states and dynamic transitions occur at four points a – d for dimensionless cycle time $T_s = 3.0$.

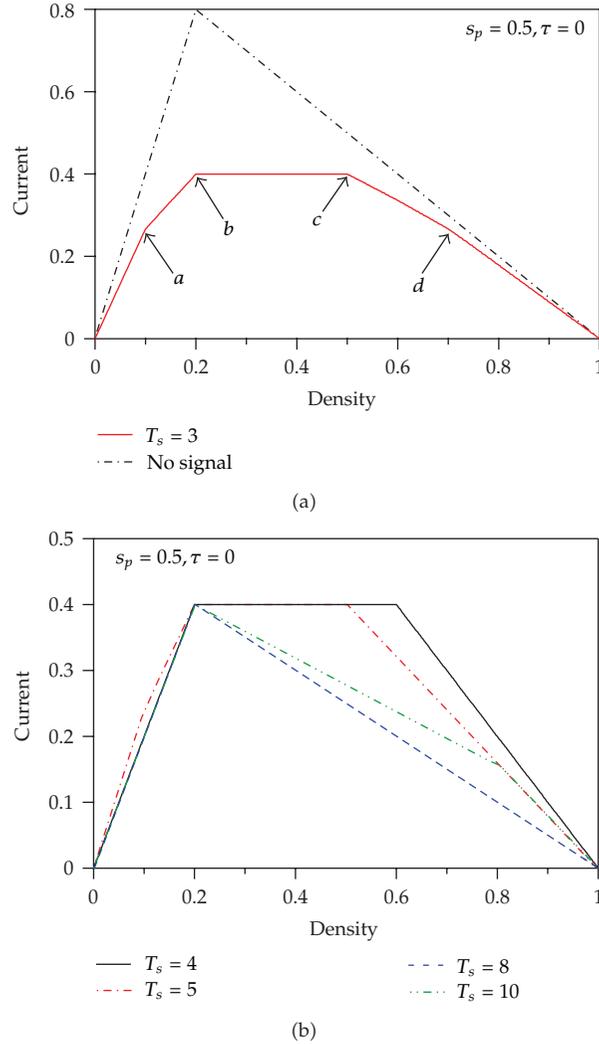


Figure 2: (a) Plots of mean current against density for $T_s = 3.0$ where the split is $s_p = 0.5$, the offset time is $\tau = 0.0$, and the road length is $L = 4000$. The solid curve indicates the current profile. For comparison, the current for the vehicular traffic with no signals is shown by the chain curve. (b) Fundamental diagrams for the dimensionless cycle times $T_s = 4.0, 5.0, 8.0, 10.0$.

We study the variation of fundamental diagram with the dimensionless cycle time. Figure 2(b) shows the fundamental diagrams for the dimensionless cycle times $T_s = 4.0, 5.0, 8.0, 10.0$. The fundamental diagram depends greatly on the dimensionless cycle time. Especially, the fundamental diagram at high values of dimensionless cycle time is definitely different from that at low value of the cycle time. With increasing the cycle time, the current profile changes from the typical trapezoid to the triangle.

We study the traffic pattern (vehicular trajectories) for various values of density at dimensionless cycle time $T_s = 3.0$. Figure 3 shows the trajectories of vehicles between $x = 400$ and $x = 600$ at various values of density for cycle time $T_s = 3.0$ where the road length is $L = 4000$, the interval between signals is $l = 40$, and the maximal velocity is $v_{\max} = 4$.

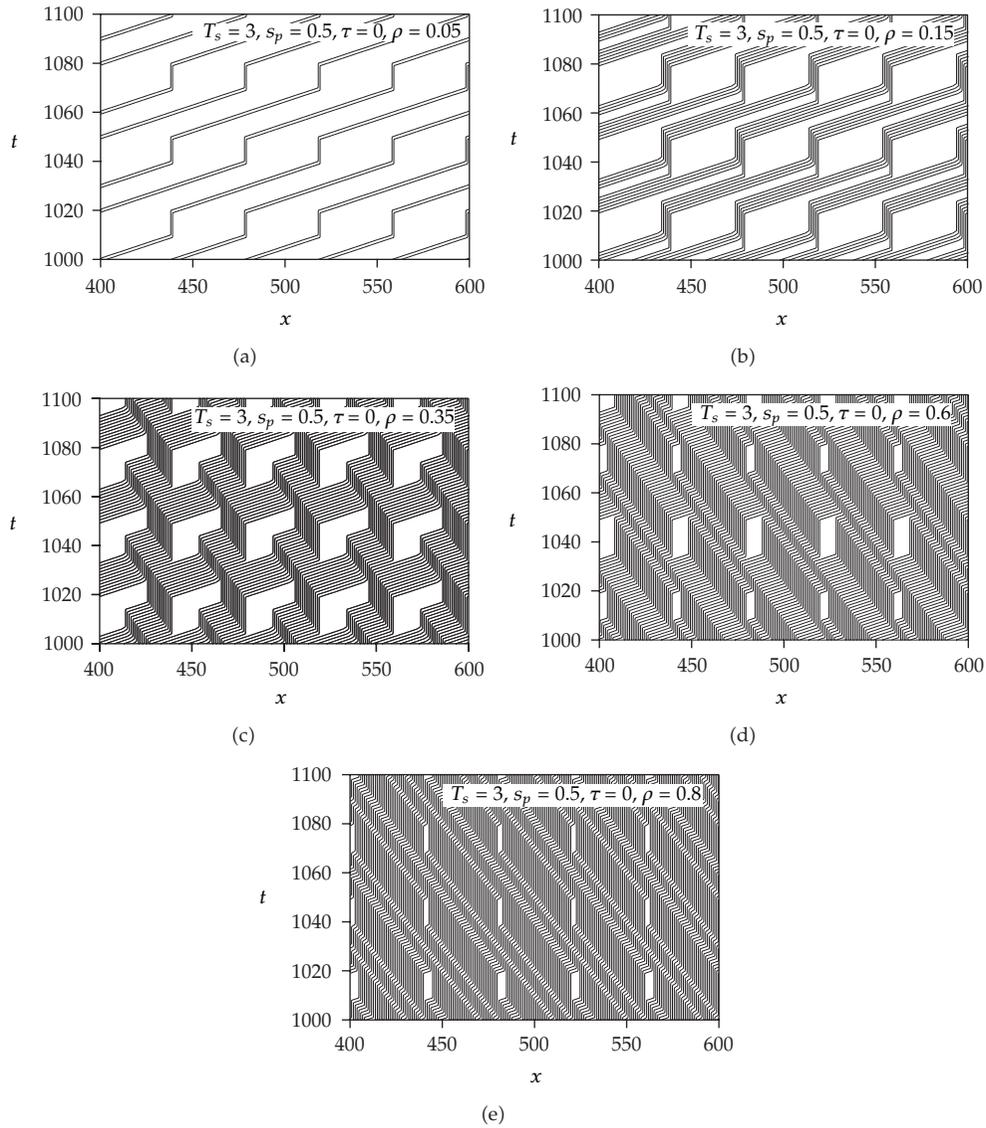


Figure 3: Trajectories of vehicles between $x = 400$ and $x = 600$ at various values of density for cycle time $T_s = 3.0$ where the road length is $L = 4000$, the interval between signals is $l = 40$, and the maximal velocity is $v_{\max} = 4$. Diagrams (a)–(e) are obtained, respectively, at densities (a) $\rho = 0.05$, (b) $\rho = 0.15$, (c) $\rho = 0.35$, (d) $\rho = 0.6$, and (e) $\rho = 0.8$ for the fundamental diagram in Figure 2(a).

Diagrams (a)–(e) are obtained, respectively, at densities (a) $\rho = 0.05$, (b) $\rho = 0.15$, (c) $\rho = 0.35$, (d) $\rho = 0.6$, and (e) $\rho = 0.8$ for the fundamental diagram in Figure 2(a). In diagram (a), a pair of two vehicles move together and stop every two signals. The traffic pattern (a) is typical for the traffic state at the region 0-a in Figure 2(a). In diagram (b), a group of six vehicles move together. However, the group breaks up into two subgroups at a signal and merges again at the next to next signal. The breakup and coalescence are repeated. The traffic pattern (b) is typical for the traffic state at the region a-b in Figure 2(a). The traffic state (a) changes to

the traffic state (b) at transition point b in Figure 2(a). In diagram (c), the stop- and go-wave is induced by stopping at the signal. The vehicles stop temporarily by the stop- and go-wave before they reach the signal. The stop- and go-wave disappears within the interval between two signals. In the traffic state (c), the traffic current saturates and keeps the maximum value. The traffic pattern (c) is typical for the traffic state at the region $b-c$ in Figure 2(a). At the transition point b , the traffic state (b) changes to the traffic state (c). In diagrams (d) and (e), the stop- and go-wave propagates backward through signals. In diagram (d), the region of the moving vehicles breaks up and coalescences, while the region of moving vehicles propagates backward with no breakup in diagram (e). At transition point c in Figure 2(a), the traffic state (c) changes to the traffic state (d). The traffic state (d) changes to the traffic state (e) at the transition point d . Thus, the traffic state changes at the transition points $a-d$. The vehicular trajectories change highly with density.

We study the dependence of the maximum current on the dimensionless cycle time. Also, we study the effect of the split s_p on the maximum current Q_{\max} and the dynamic transitions. Figure 4(a) shows the plot of the maximum currents Q_{\max} against dimensionless cycle time T_s at splits $s_p = 0.25, 0.5, 0.75$. The squares, circles, and triangles indicate, respectively, the maximum currents at $s_p = 0.25, 0.5, 0.75$. At split $s_p = 0.5$, the maximum current goes up and down with increasing dimensionless cycle time and the width decreases with increasing dimensionless cycle time. The profile displays a saw-toothed shape. When T_s approaches to an infinity, the maximum current approaches to 0.4. With increasing the split, the maximum current increases. When the split goes to infinity, the maximum currents at $s_p = 0.25, 0.5, 0.75$ approach, respectively, to 0.2, 0.4, 0.6. We rescale the profiles of the maximum current versus the dimensionless cycle time as Figure 4(b).

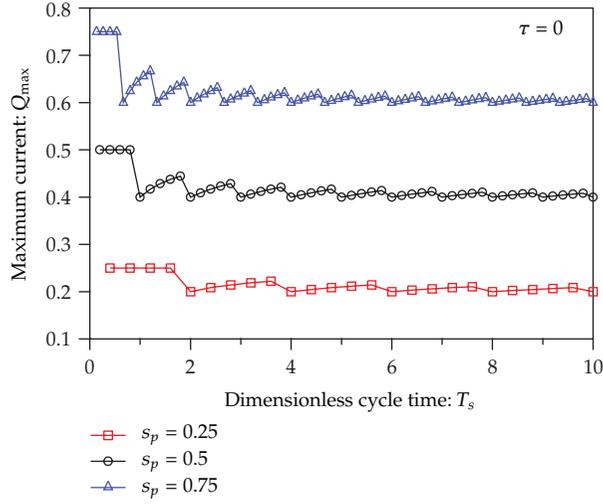
Figure 4(b) shows the plots of rescaled maximum current $Q_{\max}/2s_p$ against rescaled dimensionless cycle time $T_s s_p$ for Figure 4(a). All rescaled maximum currents collapse on a single curve. The maximum current scales as

$$Q_{\max} = 2s_p f(T_s s_p), \quad (3.1)$$

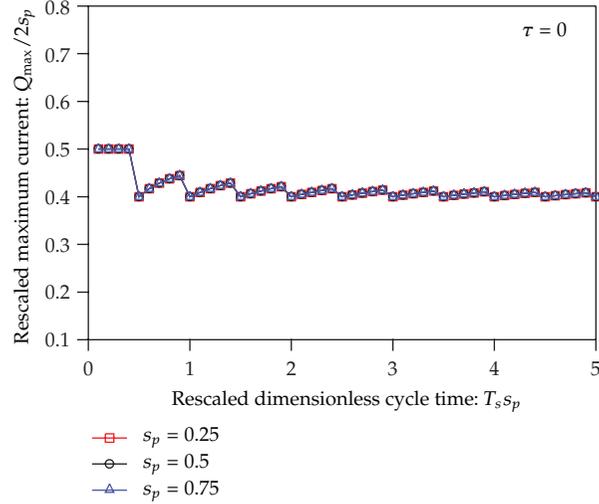
where $f(x)$ is the scaling function.

In the synchronized strategy, the maximum current is proportional to the split. Also, the dependence of the maximum current on both cycle time and split is determined only by the product $T_s s_p$. It is important how the saturated traffic disappears with density. We study the dependence of transition points b and c on the dimensionless cycle time. Also, we study the dependence of transition points b and c on the split s_p . Figure 5(a) shows the plots of the transition densities $\rho_{c,b}$ and $\rho_{c,c}$ against dimensionless cycle time at splits $s_p = 0.25, 0.5, 0.75$. The open squares, open circles, and open triangles indicate, respectively, the transition density $\rho_{c,b}$ of the dynamic transition from the undersaturated traffic to the saturated traffic at splits $s_p = 0.25, 0.5, 0.75$. The full squares, full circles, and full triangles represent, respectively, the transition density $\rho_{c,c}$ of the dynamic transition from the saturated traffic to the oversaturated traffic at splits $s_p = 0.25, 0.5, 0.75$.

At split $s_p = 0.5$, the transition point $\rho_{c,b}$ keeps constant value 0.2 for $T_s > 1.8$. The transition point $\rho_{c,c}$ varies up and down for $T_s < 4$, then decreases linearly with increasing T_s , and keeps the constant value 0.2 for $T_s > 7.8$. The transition point $\rho_{c,c}$ coalescences with the transition point $\rho_{c,b}$ at $T_s = 7.8$. The saturated traffic disappears at $T_s = 7.8$ for $s_p = 0.5$. In the result, the fundamental diagram changes from the trapezoid to the triangle at dimensionless cycle time $T_s = 7.8$. We rescale the profiles of the transition points versus



(a)



(b)

Figure 4: (a) Plot of the maximum currents Q_{\max} against dimensionless cycle time T_s at splits $s_p = 0.25, 0.5, 0.75$. The squares, circles, and triangles indicate, respectively, the maximum currents at $s_p = 0.25, 0.5, 0.75$. (b) Plots of rescaled maximum current $Q_{\max}/2s_p$ against rescaled dimensionless cycle time $T_s s_p$ for (a).

the dimensionless cycle time as Figure 5(b). Figure 5(b) shows the plots of transition points $\rho_{c,b}$ and $\rho_{c,c}$ against rescaled dimensionless cycle time $T_s s_p$ for Figure 5(a). The transition points $\rho_{c,b}$ at $s_p = 0.25, 0.5, 0.75$ collapse on a single curve. Also, the transition points $\rho_{c,c}$ at $s_p = 0.25, 0.5, 0.75$ collapse on a single curve. The transition points $\rho_{c,b}$ and $\rho_{c,c}$ scale as

$$\begin{aligned} \rho_{c,b} &= g_1(T_s s_p), \\ \rho_{c,c} &= g_2(T_s s_p), \end{aligned} \quad (3.2)$$

where $g_1(x)$ and $g_2(x)$ are the scaling functions for the two dynamic transitions.

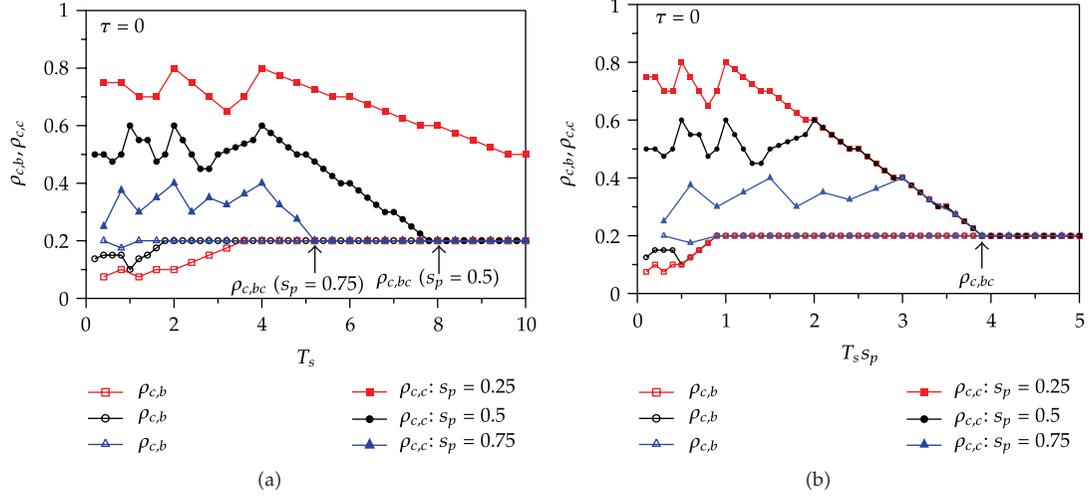
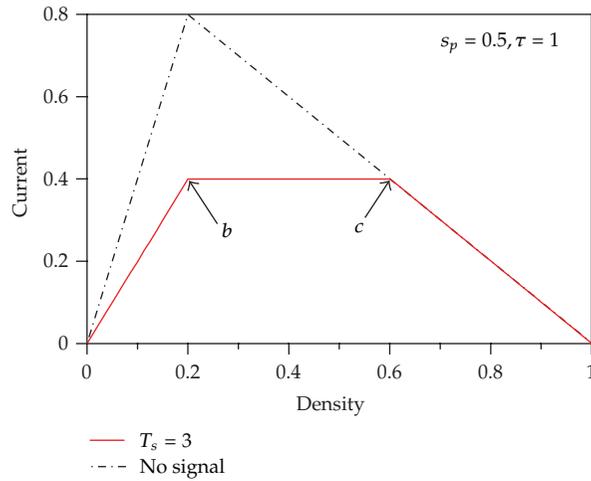


Figure 5: (a) Plots of the transition densities $\rho_{c,b}$ and $\rho_{c,c}$ against dimensionless cycle time T_s at splits $s_p = 0.25, 0.5, 0.75$. The open squares, open circles, and open triangles indicate, respectively, the transition density $\rho_{c,b}$ of the dynamic transition from the undersaturated traffic to the saturated traffic at splits $s_p = 0.25, 0.5, 0.75$. The full squares, full circles, and full triangles represent, respectively, the transition density $\rho_{c,c}$ of the dynamic transition from the saturated traffic to the oversaturated traffic at splits $s_p = 0.25, 0.5, 0.75$. (b) Plots of transition points $\rho_{c,b}$ and $\rho_{c,c}$ against rescaled dimensionless cycle time $T_s s_p$ for (a).

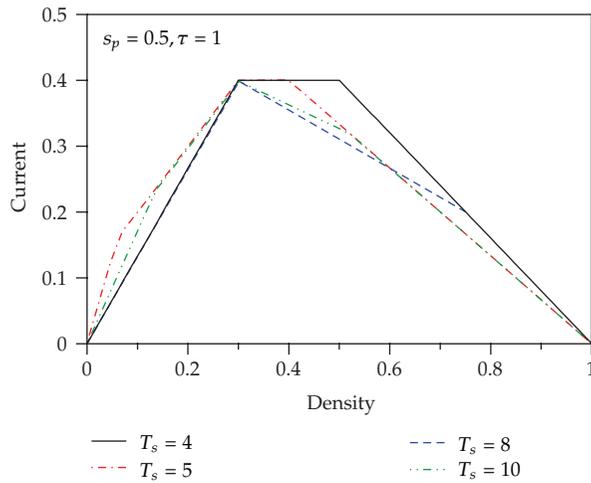
In the synchronized strategy, the transition points $\rho_{c,b}$ and $\rho_{c,c}$ depend only on the rescaled cycle time $T_s s_p$. If the product $T_s s_p$ of dimensionless cycle time T_s by split s_p is the same value, the dynamic transition between the undersaturated traffic and the saturated traffic occurs at the same value of the transition density. Also, when the product $T_s s_p$ of dimensionless cycle time T_s by split s_p is the same value, the dynamic transition between the saturated traffic and the oversaturated traffic occurs at the same value of the transition density. Thus, the dynamic transitions are governed only by the product $T_s s_p$.

3.2. Green-Wave Strategy

We study the traffic flow controlled by signals at the green-wave strategy. We investigate the effect of the offset time (phase shift difference) on the traffic flow. Signals vary from the upstream to the downstream and vice versa with constant value t_{offset} of the phase shift difference. We define the dimensionless offset time as $\tau = t_{\text{offset}} v_{\text{max}} / l$. We calculate the current (flow) versus density for various values of cycle time t_s at dimensionless offset time $\tau = 1.0$. We set the interval between signals as $l = 40$ and the maximal velocity as $v_{\text{max}} = 4$. The dimensionless cycle time is defined as $T_s = t_s v_{\text{max}} / l$. Figure 6(a) shows the plots of mean current against density at offset time $\tau = 1.0$ and $T_s = 3.0$ where the split is $s_p = 0.5$ and the road length is $L = 4000$. The solid curve indicates the current profile. For comparison, the current for the vehicular traffic with no signals is shown by the chain curve. Figure 6(a) is compared with Figure 2(a) at the synchronized strategy. The traffic states change from the undersaturated traffic, through the saturated traffic, to the oversaturated traffic. In the unsaturated traffic region between points 0–b, the current increases with density, only a state



(a)



(b)

Figure 6: (a) Plots of mean current against density at offset time and where the split is $s_p = 0.5$ and the road length is $L = 4000$. The solid curve indicates the current profile. For comparison, the current for the vehicular traffic with no signals is shown by the chain curve. (b) Fundamental diagrams for the dimensionless cycle times.

occurs, and the traffic state does not change at point a in Figure 2(a). In the saturated traffic region between points b - c , the current keeps a constant value. In the oversaturated traffic region between points c -1, the traffic state does not change at point d in Figure 2(a). Thus, there exist three distinct traffic states and dynamic transitions occur at two points b and c for dimensionless cycle time $T_s = 3.0$. Thus, the fundamental diagram in Figure 2(a) for the synchronized strategy changes to that in Figure 6(a) for $\tau = 1.0$.

We study the variation of fundamental diagram at $\tau = 1.0$ with the dimensionless cycle time. Figure 6(b) shows the fundamental diagrams at $\tau = 1.0$ for the dimensionless cycle times $T_s = 4.0, 5.0, 8.0, 10.0$. Figure 6(b) is compared with Figure 2(b) at the synchronized strategy. The fundamental diagram depends greatly on the dimensionless cycle time.

Especially, the fundamental diagram at high values of dimensionless cycle time is definitely different from that at low value of the cycle time. With increasing the cycle time, the current profile changes from the typical trapezoid to the triangle. However, the dependence of transition densities on the dimensionless cycle time is definitely different from that in Figure 2(b) at the synchronized strategy.

We study the dependence of transition points b and c on the split s_p at $\tau = 1.0$. Figure 7(a) shows the plots of the transition densities $\rho_{c,b}$ and $\rho_{c,c}$ against dimensionless cycle time at splits $s_p = 0.25, 0.5, 0.75$ at $\tau = 1.0$. The open squares, open circles, and open triangles indicate, respectively, the transition density $\rho_{c,b}$ of the dynamic transition from the undersaturated traffic to the saturated traffic at splits $s_p = 0.25, 0.5, 0.75$. The full squares, full circles, and full triangles represent, respectively, the transition density $\rho_{c,c}$ of the dynamic transition from the saturated traffic to the oversaturated traffic at splits $s_p = 0.25, 0.5, 0.75$. We rescale the profiles of the transition points versus the dimensionless cycle time as Figure 7(b). Figure 7(b) shows the plots of rescaled transition points $\rho_{c,b} - (1 - s_p)/5$ and $\rho_{c,c} - (1 - s_p)/5$ against rescaled dimensionless cycle time $T_s s_p$ for Figure 7(a). The rescaled transition points scale as those at the synchronized strategy in Figure 5(b). The transition points $\rho_{c,b}$ at $s_p = 0.25, 0.5, 0.75$ collapse on a single curve. Also, the transition points $\rho_{c,c}$ at $s_p = 0.25, 0.5, 0.75$ collapse on a single curve. The transition points $\rho_{c,b}$ and $\rho_{c,c}$ scale as

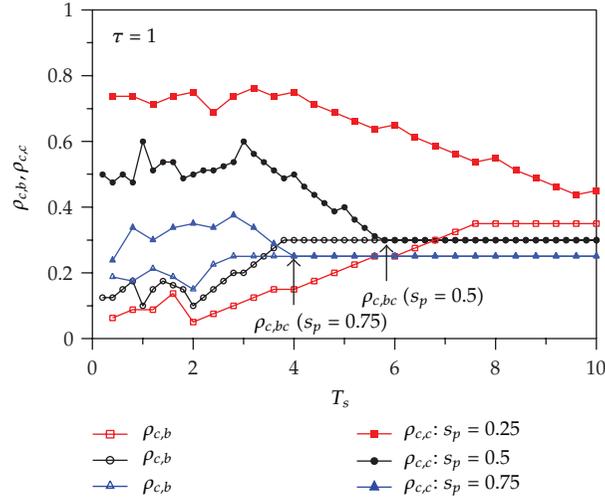
$$\begin{aligned}\rho_{c,b} &= \frac{1 - s_p}{5} + g_3(T_s s_p), \\ \rho_{c,c} &= \frac{1 - s_p}{5} + g_4(T_s s_p),\end{aligned}\tag{3.3}$$

where $g_3(x)$ and $g_4(x)$ are the scaling functions for the two dynamic transitions.

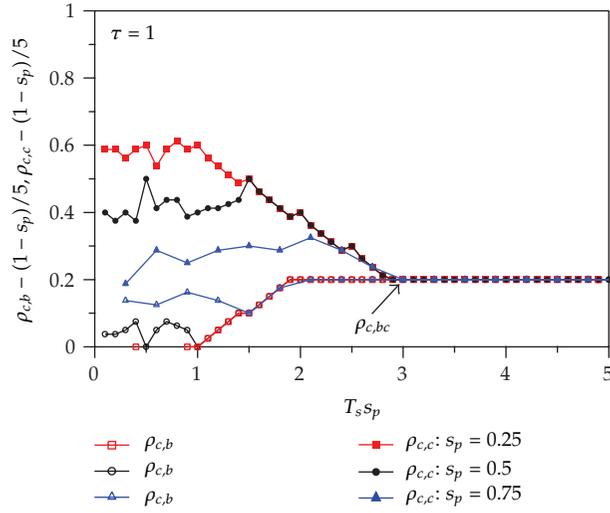
All rescaled transition points keep 0.2 for $T_s s_p \geq 3.0$. The fundamental diagram changes from the trapezoid to the triangle at point $\rho_{c,bc}$ ($T_s s_p = 3.0$).

We calculate the current (flow) versus density for various values of cycle time t_s at a negative value of dimensionless offset time $\tau = -1.0$. Figure 8 shows the fundamental diagrams at $\tau = -1.0$ for the dimensionless cycle times $T_s = 4.0, 5.0, 8.0, 10.0$. Figure 8 is compared with Figure 2(b) at the synchronized strategy and Figure 6(b) at $\tau = 1.0$. The fundamental diagram depends greatly on the dimensionless offset time τ . The fundamental diagram at $\tau = -1.0$ is definitely different from that at $\tau = 0.0$ and that at $\tau = 1.0$. With increasing the cycle time, the current profile changes from the typical trapezoid to the triangle. However, the dependence of transition densities on the dimensionless cycle time is definitely different from that in Figure 6(b) at $\tau = 1.0$.

We study the dependence of transition points b and c on the split s_p at $\tau = -1.0$. Figure 9(a) shows the plots of the transition densities $\rho_{c,b}$ and $\rho_{c,c}$ against dimensionless cycle time at splits $s_p = 0.25, 0.5, 0.75$ at $\tau = -1.0$. We rescale the profiles of the transition points versus the dimensionless cycle time as Figure 9(b). Figure 9(b) shows the plots of rescaled transition points $\rho_{c,b} + (1 - s_p)/5$ and $\rho_{c,c} + (1 - s_p)/5$ against rescaled dimensionless cycle time $T_s s_p$ for Figure 9(a). The transition points $\rho_{c,b}$ at $s_p = 0.25, 0.5, 0.75$ collapse on a single



(a)



(b)

Figure 7: (a) Plots of the transition densities $\rho_{c,b}$ and $\rho_{c,c}$ against dimensionless cycle time at splits $s_p = 0.25, 0.5, 0.75$ at $\tau = 1.0$. (b) Plots of rescaled transition points $\rho_{c,b} - (1 - s_p)/5$ and $\rho_{c,c} - (1 - s_p)/5$ against rescaled dimensionless cycle time $T_s s_p$ for (a).

curve. Also, the transition points $\rho_{c,c}$ at $s_p = 0.25, 0.5, 0.75$ collapse on a single curve. The transition points $\rho_{c,b}$ and $\rho_{c,c}$ scale as

$$\begin{aligned}\rho_{c,b} &= -\frac{1-s_p}{5} + g_5(T_s s_p), \\ \rho_{c,c} &= -\frac{1-s_p}{5} + g_6(T_s s_p),\end{aligned}\tag{3.4}$$

where $g_5(x)$ and $g_6(x)$ are the scaling functions for the two dynamic transitions.

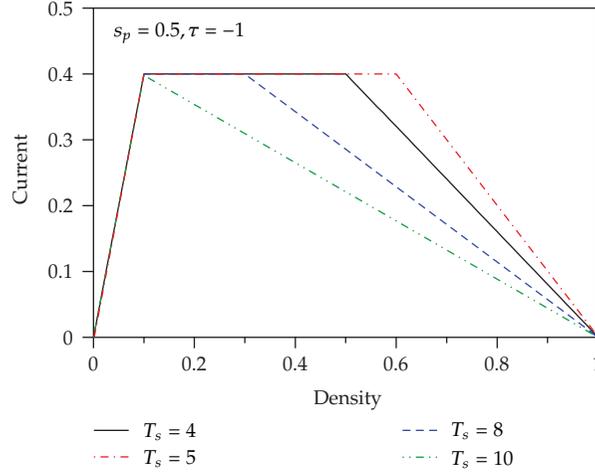


Figure 8: Fundamental diagrams at $\tau = -1.0$ for the dimensionless cycle times $T_s = 4.0, 5.0, 8.0, 10.0$.

We study the dependence of the transition points $\rho_{c,b}$ and $\rho_{c,c}$ on the dimensionless cycle time for various values of the offset time τ . Figure 10(a) shows the plots of transition densities $\rho_{c,b}$ and $\rho_{c,c}$ against the dimensionless cycle time for $\tau = 0.0, 1.0, 2.0, 3.0$ at split $s_p = 0.5$. Figure 10(b) shows the plot of the rescaled transition density $\rho_{c,b} - \tau(1 - s_p)/5$ against the rescaled cycle time $T_s s_p - \tau$ for Figure 10(a). The transition points $\rho_{c,b}$ for $\tau = 0.0, 1.0, 2.0, 3.0$ collapse on a single curve. Figure 10(c) shows the plot of the rescaled transition density $\rho_{c,c} - \tau(1 - s_p)/5$ against the rescaled cycle time $T_s s_p + \tau$ for Figure 10(a). The transition points $\rho_{c,c}$ for $\tau = 0.0, 1.0, 2.0, 3.0$ approach to a single curve. The transition points $\rho_{c,b}$ and $\rho_{c,c}$ scale for $0 \leq \tau \leq 3.0$ as

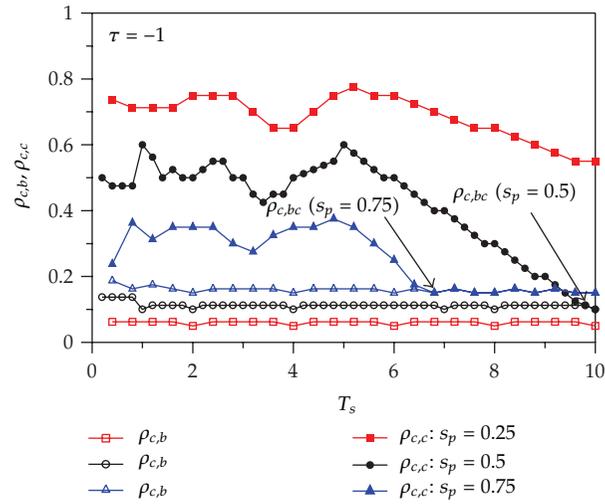
$$\rho_{c,b} = \frac{\tau(1 - s_p)}{5} + g_b(T_s s_p - \tau), \quad (3.5)$$

$$\rho_{c,c} = \frac{\tau(1 - s_p)}{5} + g_c(T_s s_p + \tau), \quad (3.6)$$

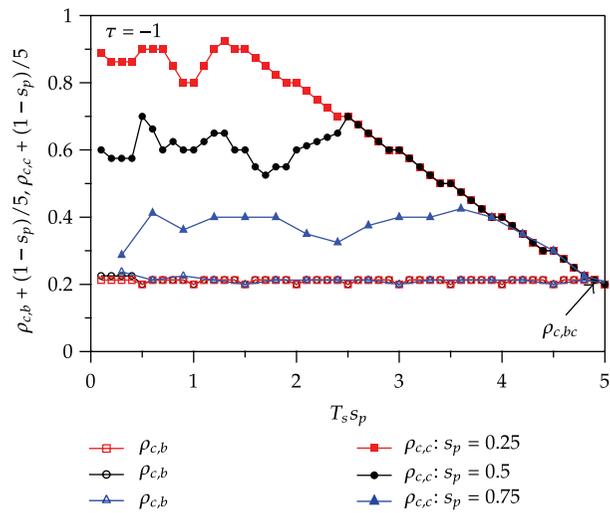
where $g_b(x)$ and $g_c(x)$ are the scaling functions for the two dynamic transitions.

The transition points $\rho_{c,b}$ and $\rho_{c,c}$ scale for $-1.0 \leq \tau \leq 0.0$ as (3.5) and (3.6). The scaling forms (3.5) and (3.6) of transition points $\rho_{c,b}$ and $\rho_{c,c}$ hold only for $-1.0 \leq \tau \leq 3.0$. The scaling forms are not satisfied for the values of the dimensionless offset time except for $-1.0 \leq \tau \leq 3.0$ but the profiles of the transition points versus the dimensionless cycle time deviate from the scaling forms for $\tau < -1.0$ and $\tau > 3.0$.

Thus, we obtain the following finding. The transition point from the undersaturated traffic to the saturated traffic scales as (3.5) for any values of split and offset time. Also, the transition point from the saturated traffic to the oversaturated traffic scales as (3.6) for any values of split and offset time.



(a)



(b)

Figure 9: (a) Plots of the transition densities $\rho_{c,b}$ and $\rho_{c,c}$ against dimensionless cycle time at splits $s_p = 0.25, 0.5, 0.75$ at $\tau = -1.0$. (b) Plots of rescaled transition points $\rho_{c,b} + (1 - s_p)/5$ and $\rho_{c,c} + (1 - s_p)/5$ against rescaled dimensionless cycle time $T_s s_p$ for (a).

4. Summary

The deterministic cellular automaton model was presented for vehicular traffic through the series of traffic signals. The CA model is described by the difference equation. We have studied the effect of both vehicular density and signal's characteristics on dynamic behavior of vehicles by using the CA model. The fundamental diagram was derived for various values of the cycle time. The effect of the dimensionless cycle time on the fundamental diagram was clarified. It was found that the fundamental diagram changes from the typical trapezoid to the triangle with increasing the dimensionless cycle time. It was shown that the fundamental

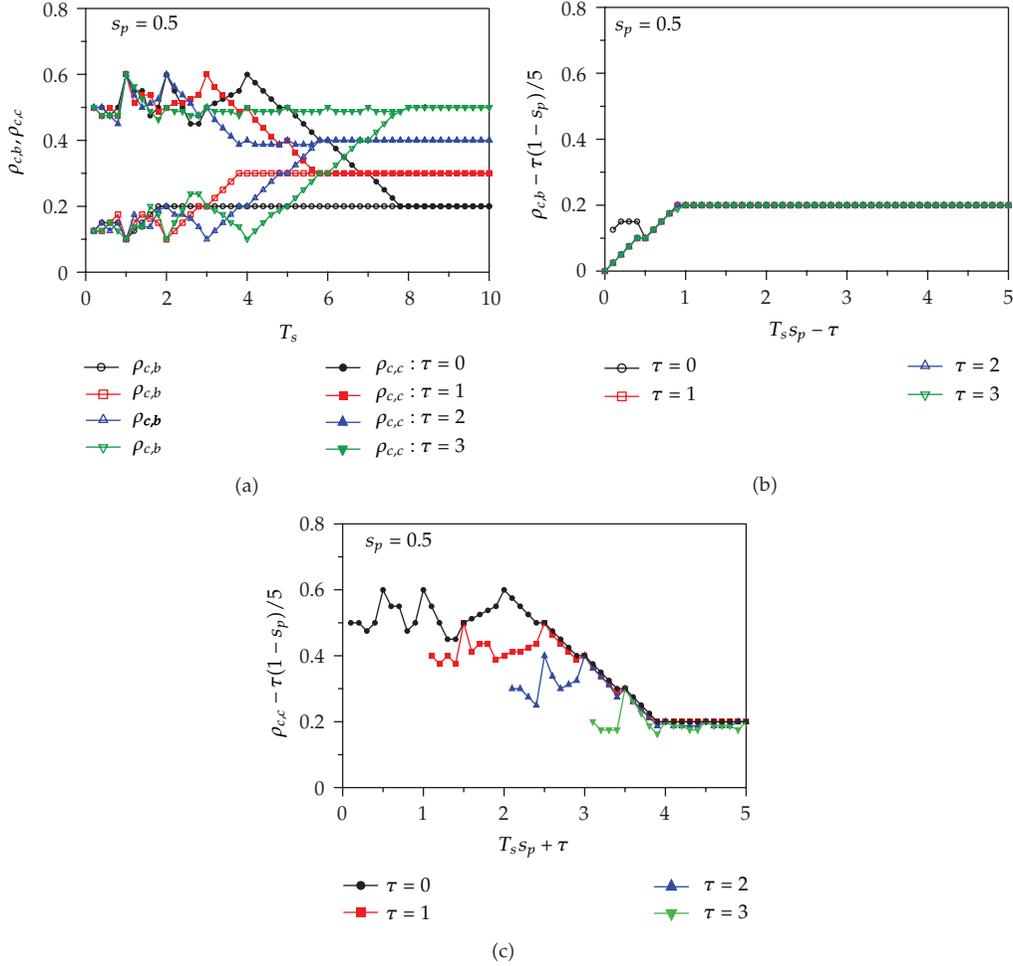


Figure 10: (a) Plots of transition densities $\rho_{c,b}$ and $\rho_{c,c}$ against the dimensionless cycle time for $\tau = 0, 1, 2, 3$ at split $s_p = 0.5$. (b) Plot of the rescaled transition density $\rho_{c,b} - \tau(1 - s_p)/5$ against the rescaled cycle time $T_s s_p - \tau$ for (a). (c) Plot of the rescaled transition density $\rho_{c,c} - \tau(1 - s_p)/5$ against the rescaled cycle time $T_s s_p + \tau$ for (a).

diagram also depends highly on the split and the dimensionless offset time. It was found that the transition densities from the undersaturated traffic to the saturated traffic and from the saturated traffic to the oversaturated traffic scale as (3.5) and (3.6) for various values of both split and offset time.

The deterministic CA model proposed by this paper will be useful for studying the traffic flow through a series of signals controlled by the cycle time, the split, and the offset time because it is described by the difference equation. The study for the dependence of the dynamic transitions on the cycle time, the split, and the offset time is the first. Especially, the change of the fundamental diagram from the trapezoid to the triangle will be interesting.

References

- [1] T. Nagatani, "The physics of traffic jams," *Reports on Progress in Physics*, vol. 65, no. 9, article 1331, 2002.
- [2] D. Helbing, "Traffic and related self-driven many-particle systems," *Reviews of Modern Physics*, vol. 73, no. 4, pp. 1067–1141, 2001.
- [3] D. Chowdhury, L. Santen, and A. Schadscheider, "Statistical physics of vehicular traffic and some related systems," *Physics Reports*, vol. 329, no. 4–6, pp. 199–329, 2000.
- [4] B. S. Kerner, *The Physics of Traffic*, Springer, Heidelberg, Germany, 2004.
- [5] A. Kirchner and A. Schadschneider, "Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics," *Physica A*, vol. 312, no. 1-2, pp. 260–276, 2002.
- [6] 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.
- [7] G. H. Peng, X. H. Cai, C. Q. Liu, and B. F. Cao, "A new lattice model of traffic flow with the consideration of the driver's forecast effects," *Physics Letters A*, vol. 375, no. 22, pp. 2153–2157, 2011.
- [8] T. Nagatani, "Chaotic jam and phase transition in traffic flow with passing," *Physical Review E*, vol. 60, no. 2, pp. 1535–1541, 1999.
- [9] L. A. Safonov, E. Tomer, V. V. Strygin, Y. Ashkenazy, and S. Havlin, "Multifractal chaotic attractors in a system of delay-differential equations modeling road traffic," *Chaos*, vol. 12, no. 4, pp. 1006–1014, 2002.
- [10] E. Brockfeld, R. Barlovic, A. Schadschneider, and M. Schreckenberg, "Optimizing traffic lights in a cellular automaton model for city traffic," *Physical Review E*, vol. 64, no. 5, Article ID 056132, 12 pages, 2001.
- [11] D. W. Huang and W. N. Huang, "Traffic signal synchronization," *Physical Review E*, vol. 67, no. 5, Article ID 056124, 7 pages, 2003.
- [12] M. Sasaki and T. Nagatani, "Transition and saturation of traffic flow controlled by traffic lights," *Physica A*, vol. 325, no. 3-4, pp. 531–546, 2003.
- [13] S. Lammer and D. Helbing, "Self-control of traffic lights and vehicle flows in urban road networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, Article ID P04019, 2008.
- [14] B. A. Toledo, V. Munoz, J. Rogan, and C. Tenreiro, "Modeling traffic through a sequence of traffic lights," *Physical Review E*, vol. 70, no. 1, Article ID 016107, 6 pages, 2004.
- [15] B. A. Toledo, E. Cerda, J. Rogan et al., "Universal and nonuniversal features in a model of city traffic," *Physical Review E*, vol. 75, no. 2, Article ID 026108, 10 pages, 2007.
- [16] T. Nagatani, "Self-similar behavior of a single vehicle through periodic traffic lights," *Physica A*, vol. 347, pp. 673–682, 2005.
- [17] T. Nagatani, "Control of vehicular traffic through a sequence of traffic lights positioned with disordered interval," *Physica A*, vol. 368, no. 2, pp. 560–566, 2006.
- [18] T. Nagatani, "Clustering and maximal flow in vehicular traffic through a sequence of traffic lights," *Physica A*, vol. 377, no. 2, pp. 651–660, 2007.
- [19] T. Nagatani, "Effect of irregularity on vehicular traffic through a sequence of traffic lights," *Physica A*, vol. 387, no. 7, pp. 1637–1647, 2008.
- [20] T. Nagatani, "Traffic states and fundamental diagram in cellular automaton model of vehicular traffic controlled by signals," *Physica A*, vol. 388, no. 8, pp. 1673–1681, 2009.
- [21] C. Chen, J. Chen, and X. Guo, "Influences of overtaking on two-lane traffic with signals," *Physica A*, vol. 389, no. 1, pp. 141–148, 2010.
- [22] M. Fukui and Y. Ishibashi, "Traffic flow in 1D cellular automaton model including cars moving with high speed," *Journal of the Physical Society of Japan*, vol. 65, pp. 1868–1870, 1996.

Research Article

Track Irregularity Time Series Analysis and Trend Forecasting

Jia Chaolong,¹ Xu Weixiang,² Wang Futian,¹ and Wang Hanning¹

¹ State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

² School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Jia Chaolong, jiachaolong@bjtu.edu.cn

Received 27 August 2012; Revised 27 October 2012; Accepted 27 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Jia Chaolong 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 combination of linear and nonlinear methods is widely used in the prediction of time series data. This paper analyzes track irregularity time series data by using gray incidence degree models and methods of data transformation, trying to find the connotative relationship between the time series data. In this paper, GM (1, 1) is based on first-order, single variable linear differential equations; after an adaptive improvement and error correction, it is used to predict the long-term changing trend of track irregularity at a fixed measuring point; the stochastic linear AR, Kalman filtering model, and artificial neural network model are applied to predict the short-term changing trend of track irregularity at unit section. Both long-term and short-term changes prove that the model is effective and can achieve the expected accuracy.

1. Introduction

Track irregularity is a serious threat to the safety of train operation. Track irregularity data includes environmental variables (gauge, longitudinal level, cross level, alignment, and twist) and effective variables (vertical acceleration and horizontal acceleration). The developing and changing process of the track irregularity state is random, which cannot be defined by identified function. Generally, it can be researched with the combination of probability theory and analysis method within a certain range. Nowadays, most studies focus on the overall indicators which evaluate the changes of the track's state, but a few studies focus on the changes of specific geometric parameters' changes and the laws behind them. This is a basic difficulty.

Linear and nonlinear methods are two groups of models employed to estimate time series. DENG Julong [1] proposed the gray system theory in 1982. Gray system theory has

been widely applied to the field of controlling, forecasting, and decision making, and the GM (1,1) model is its core essence. G. Liu and Yu [2] studied the main factors that could affect MLF generation by using the method of gray correlation coefficient. Marcellino et al. [3] and Ding et al. [4] studied the autoregressive model (AR) to forecast macroeconomic time series and parameter estimation problems. AR is a main model of random process, which can only reflect the target through historical values of the time series, without being constrained by the mutually independent variables, eliminating the difficulties caused by independent variables selection in ordinary regression prediction and multicollinearity, and so forth. Kalman [5] proposed Kalman filter model in 1960. In the study of Feil et al. [6] and Kandepu et al. [7], Kalman model was applied to monitoring process transitions and nonlinear state estimation. Rumelhart and McClelland [8] studied the neural network years ago. Balestrassi et al. [9] studied neural network's training for nonlinear time series forecasting. Khashei et al. [10] studied artificial neural networks in hybrid models. The hybrid method is widely used on predicting time series predictions now. Zhang [11] proposed to take advantage of the unique strengths of ARIMA and ANN models in linear and nonlinear modeling. H. Liu et al. [12] studied hybrid methods in the prediction of wind speed based on time series, artificial neural networks (ANNs), and Kalman filter (KF). Areekul et al. [13] studied hybrid methodology which combined both autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) models to predict short-term electricity prices. Khashei et al. [14] and Khashei and Bijari [15] proposed hybrid method that could yield more accurate results with incomplete data sets based on the basic concepts of ANNs and fuzzy; he also proposed hybrid model of artificial neural networks by using autoregressive integrated moving average (ARIMA) models in order to yield a more accurate forecasting model compared to artificial neural networks. Aladag et al. [16] proposed a hybrid approach combined with Elman's Recurrent Neural Networks (ERNNs) and ARIMA models and applied the approach to Canadian Lynx data. In practical prediction, research methods are often composed of two types of models.

In this paper, three aspects are studied on trends of track cross level state changes. First, it analyzes track irregularity time series data and tries to find the connotative relationships between time series data with the application of seven gray incidence degree theories; secondly, it predicts long-term track level changes at fixed measuring point; finally, it predicts changes of tracks over time at unit section in short term. This paper modifies and corrects the inadequacies in the GM (1,1) model, which can only reflect the state of development of the general trend other than reflect cycle and random variation of the changes of track level at the fixed measuring point. The accuracy of fitting and forecasting can be greatly improved. In terms of unit section track state study, this paper uses random linear AR model and Kalman filtering model to analyze track state over time as well as to predict its future state. By combining the above studies, we can see the statistical laws of track state changes in the long and short term and can forecast the future state of the track.

2. Data Analysis

2.1. Analysis of Track Irregularity Data

The idea of time series analysis has been applied in many areas of research, such as the relationship of following speed and spacing with driving time in driver's safety-related approaching behavior [17, 18]. In track irregularity time series studies, the continuity of tracks

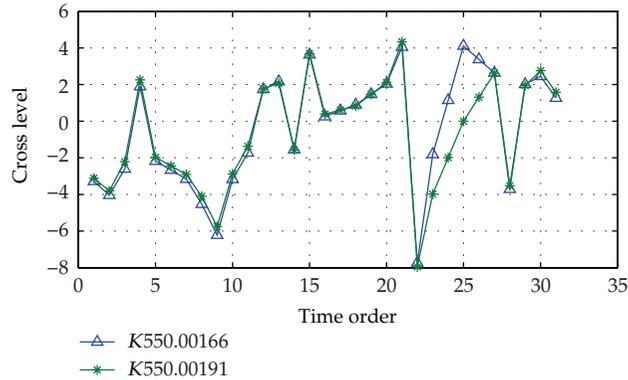


Figure 1: Comparison of track cross level values at K550.00166 and K550.00191 mileage points.

Table 1: Seven incidence degrees between cross level time series and reference cross level time series.

	DID	AID	IAID	TID	SID	FODID	SODIG
(1, R)	0.9753	0.9774	0.5032	0.3758	0.9347	0.5811	0.1399
(2, R)	0.9779	0.9767	0.4867	0.3704	0.8977	0.5144	0.1874
(3, R)	0.6299	0.7450	0.5058	0.3897	0.8143	-0.3925	-0.1512
(4, R)	0.9707	0.9768	0.4604	0.3988	0.8718	0.5400	0.1737
(5, R)	0.9743	0.9768	0.4702	0.3790	0.8911	0.5150	0.1496
...

The relationship of cross level data between adjacent hours is shown in Figure 2.

leads to a great similarity between two random time series data obtained at two adjacent inspection points. The comparison of track cross level values at K550.00166 and K550.00191 mileage points is shown in Figure 1.

It can be seen through Figure 1 that data obtained between the two adjacent measuring points shares high similarity. There is a great inconsistency during the 23th, 24th, 25th, and 26th inspection at the two adjacent measuring points. It shows great changes on track state during this time period.

In terms of the complicity of the relationships of time series curves, it is not easy to find a standard or a fixed formula to indicate the time series curve, but it can only give a complex evaluation on the changes and a developing tendency of the time series data. As a result, this paper analyzes and compares seven incidence degree algorithms. Certain relationships exist between track irregularity time series. Seven incidence degree [19] formulas include displacement incidence degree (DID), absolute incidence degree (AID), improved absolute incidence degree (IAID), T incidence degree (TID), slide incidence degree (SID), first difference incidence degree (FODID), and second difference incidence degree (SODIG). These seven incidence degree are used to reflect the corelationship between time series curves. Table 1 shows, respectively, seven incidence degrees between actual cross level time series and reference cross level time series.

It can be found from Figure 2 that changes on two adjacent cross level irregularity state data show the linear trend, with an approximated slope of 1. If there is a large deviation from the slope, it will illustrate that the two adjacent inspection data on track cross level state have been changed greatly and are in need of special attention.

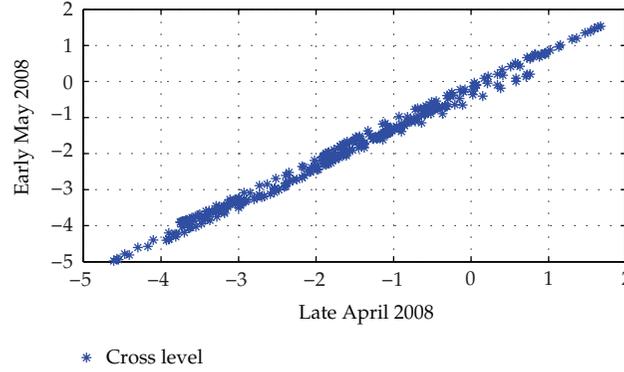


Figure 2: Scatter diagrams of detection data during the last ten-day period of April and the first ten-day period of May, 2008.

2.2. Analysis of Track Irregularity Time Series

Track inspection data refers to the data obtained within a roughly fixed time interval (a half month), which is generated from geometry state detection along the mileage range of railway line. The time sequence of track geometry state changes with the following characteristics.

(1) Data Elements of Original Time Series Is a Data Set

In the study of variation law of detection data, each detection data on a certain unit of section area is considered as a data unit. Data sequence consisted of data unit within a certain time frame is the object of study, forming a time series. Original time series data is described as follows:

$$T = \left\{ \left(x^{(1)}_{t_1}, x^{(2)}_{t_1}, \dots, x^{(n)}_{t_1} \right), \left(x^{(1)}_{t_2}, x^{(2)}_{t_2}, \dots, x^{(n)}_{t_2} \right), \dots, \left(x^{(1)}_{t_m}, x^{(2)}_{t_m}, \dots, x^{(n)}_{t_m} \right) \right\}. \quad (2.1)$$

In the formula, $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ is the prediction data set at the unit section, constituting a time series of data units, $x_{t_i} = \{t_i, m, g_i, ld_i, cl_i, cr_i, al_i, ar_i, tw_i\}$, t_i is the time point in time series, $i = 1, 2, \dots, m$, m is mileage, g_i is gauge, ld_i is longitudinal level, cl_i is cross level (L), cr_i is cross level (R), al_i is alignment (L), ar_i is alignment (R), and tw_i is twist.

(2) Data Transformation Is Necessary

Since each data unit is not a single data, but a data set of union section, rather than, therefore, it is necessary to transform processing in order to form data which can reflect the real characters of this section geometry state at t_i :

$$T = \{X_{t_1}, X_{t_2}, \dots, X_{t_i}, \dots, X_{t_m}\},$$

$$X_{t_i} = f\left(x^{(1)}_{t_i}, x^{(2)}_{t_i}, \dots, x^{(n)}_{t_i}\right) = \left[\frac{1}{n} \left(x^{(1)2}_{t_i} + x^{(2)2}_{t_i} + \dots + x^{(n)2}_{t_i} \right) \right]^{1/2}. \quad (2.2)$$

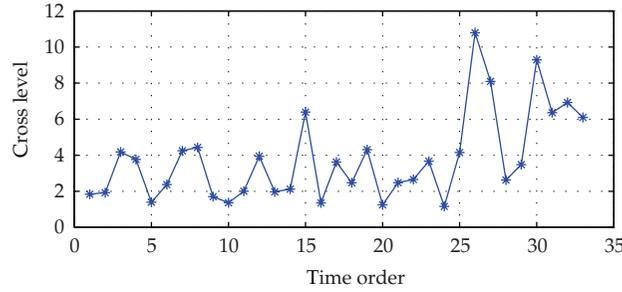


Figure 3: Changes of time series data at unit section.

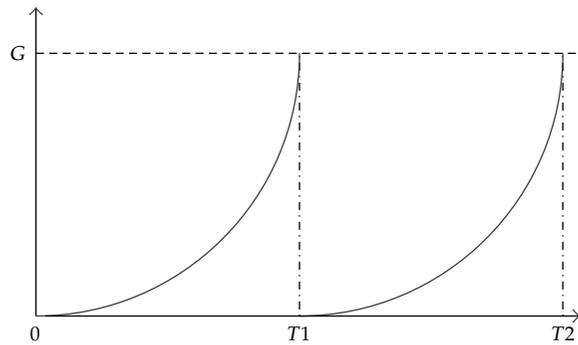


Figure 4: Cyclical trend of track state condition change.

In the formula, $f(x^{(1)}_{t_i}, x^{(2)}_{t_i}, \dots, x^{(n)}_{t_i})$ reflects the characteristics of the entire section of the track geometry at t_i . After transformation, changes of time series data at unit section are shown in Figure 3.

(3) Time Series Data Are Small Data Sets

In order to keep track status in good condition and to ensure operation safety, maintenance at regular intervals is needed as the track state changes. Only data from two maintenance operations can be seen as the objects of the study as well as time series data. It also means that this is a small data set within a short timespan. We need to find an effective forecasting method to realize our research goal even though historical data is limited.

As shown in Figure 4, G refers to track geometry state changes (deterioration) limits and $T1$ and $T2$ refer to the exact time that maintenance and repair operations occurred within. It shows a cyclical changing trend of track state conditions.

(4) Data Selecting

In this paper, track irregularity data by track inspection car in the experimentation is provided by State key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. The cross level irregularity data is selected as the object of this research. The research selects the Beijing-Kowloon upline, the K550 + 000 to K550 + 075 mileage ranging from the second track inspection in late February 2008 to the second track inspection in late May 2009, a total of

31 inspection data as data object, each of which contains 300 cross level values and each data array contains 300 elements.

3. Medium and Long-Term Track State Change Models

3.1. Improvement of GM (1, 1) Model and Prediction

GM (n, h) stands for a grey model of “ h ” variable expressed by n order differential equations. Generally speaking, when we make benefit analysis and production forecast in the field of economy and agroecology, we only work on a variable—a result, by then $h = 1$. But when n is too large, it will be too difficult to calculate; thus model GM (1, 1) will be commonly used. G represents gray, M represents model, and GM (1, 1) stands for first-order, single variable grey model [1], which can be represented by the linear differential equation (3.1). GM (1, 1) model is usually used to predict growth trend sequences with power exponent and usually has better accuracy in prediction. However, in reality, in most cases, the series data does not show the exponentially growth trend, and generally they are outliers, which limit the range of applications and fields of GM (1, 1). Thus, the model needs to be improved on the pretreatment of the raw sequence, so that it can expand the extent to which the model can be applied.

Track cross level irregularity data $x^{(0)}$ is a data set at a fixed measuring point, which fluctuates along mileage with zero values, and the data itself is not monotonic. The methods of the improved model are as follows: first, the fluctuating value of data is changed to zero by translating, and then a fixed positive constant is added to each data, so that the new time series data are positive. Next, smooth the new time series data using a power function $[x^{(0)} + I]^\alpha$. The result of new time series data weakens the impact of outliers on the fitted data. In this paper, the positive regular value I is selected as the integer value of two times of the maximum absolute value among all original series of data elements, that is, $I = \text{int}(2 \cdot \max |x^{(0)}|)$.

According to the analysis of the track cross level sequence of raw data, we find that $\text{int}(2 \cdot \max |x^{(0)}|) = 15$. So we select 15 as the positive constant value. According to the degree of the dispersion of the newly constructed data, α , which ranges from 0 to 1, can be determined. Combined with data characteristics of cross level irregularity, we set $\alpha = 0.2$. Reconstruct the original series $x^{(0)}$ by applying new methods of constructing series and then get the new series $\hat{x}^{(0)}$, adding up, AGO sequence $x^{(1)}$ is constructed:

$$\frac{dx^{(1)}}{dt} + \alpha x^{(1)} = u. \quad (3.1)$$

When $dx^{(1)}/dt$ is on a point value $[k, k + 1]$, $dx^{(1)}/dt$ approximation taken in the point $k + 1$, that is,

$$\left. \frac{dx^{(1)}}{dt} \right|_{t=k+1} \approx x^{(1)}(k + 1) - x^{(1)}(k) = \hat{x}^{(0)}(k + 1). \quad (3.2)$$

After transformation, let us solve differential equations

$$\hat{x}^{(0)}(k + 1) = a \left[-\frac{1}{2} \left(x^{(1)}(k + 1) + x^{(1)}(k) \right) \right] + u, \quad k = 1, 2, \dots, n. \quad (3.3)$$

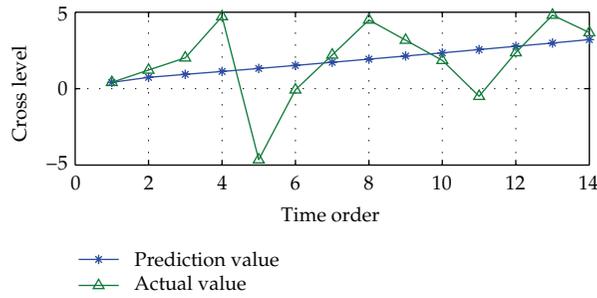


Figure 5: Comparison of actual value and prediction value at a fixed measuring point.

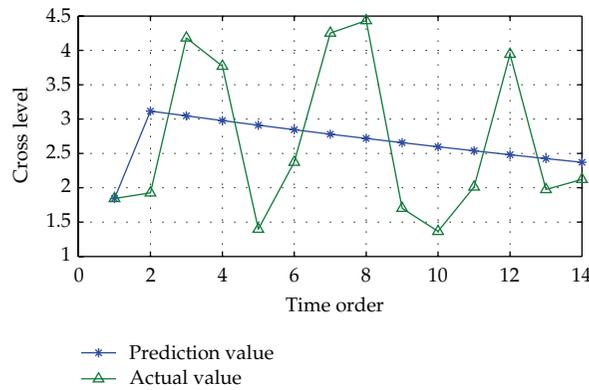


Figure 6: Comparison of actual state and predictive state at unit section.

Then we can obtain the coefficients a and u of the regression curve according to the least squares method. The expression resulted from the solutions of

$$x^{(1)}(k+1) = \left[\hat{x}^{(0)}(1) - \frac{u}{a} \right] e^{-ak} + \frac{u}{a}. \tag{3.4}$$

Next, take the values of a and u into (3.4), we can obtain GM (1, 1) prediction model of the track cross level state changes. Because the calculation data is a data array which is added up based on a fixed value; the final predictive value of the expression is as follows:

$$x^{(0)}(k) = \left(x^{(1)}(k) - x^{(1)}(k-1) \right)^{1/a} - 15. \tag{3.5}$$

With the application of the formula (3.5), we can predict the law of trend on historic track cross level data. The trend curve and actual curve fitting is shown in Figure 5.

When gray model GM (1,1) is used to forecast time series data that after transformation (see Section 2.2) at the unit section, the predictive result is shown in Figure 6.

It can be found from Figures 5 and 6 that the GM (1, 1) predictive value curve is smooth and there is a larger deviation between the predicted values and actual values; so it can only

reflect the overall trend, but cannot reflect the characteristics of the cyclical changes and random fluctuations and cannot be applied to forecast track state. Therefore, the GM (1,1) model needs amended residuals to meet the forecasting requirements.

3.2. Gray Model GM (1,1) with Residual Modification

Since the residuals are large, there will be a great inaccuracy in GM (1,1) when predicting the actual track state change trends. So we cannot predict the medium and long-term track state changes. In this paper, a method based on the trigonometric residual modification was presented to improve the predictive accuracy.

Time series of the track geometry state changes has cyclical characteristics according to the analysis of the historical changing trend of cross level. We find that trigonometric function has obvious cyclical features. In this paper, trigonometric function is used to correct residuals of the prediction model. Here, the residual refers to the actual value minus the predicted value, that is, $(\hat{x}^{(0)}(k))^{1/\alpha} - (\tilde{x}^{(0)}(k))^{1/\alpha}$. Set

$$f = \left(\hat{x}^{(0)}(k)\right)^{1/\alpha} - \left(\tilde{x}^{(0)}(k)\right)^{1/\alpha} = a_0 + A \sin(\omega(t-1)). \quad (3.6)$$

In the formula, A is the amplitude of wave mode, $\omega = 2\pi/T$, T is the cycle, t is the inspection time interval sequence. Because of $x^{(0)} = x^{(1)}$, so $a_0 = 0$. One has

$$A = \frac{2}{n} \sum_{k=1}^n A_i = \frac{2}{n} \sum_{k=1}^n \left| \left(\hat{x}^{(0)}(k)\right)^{1/\alpha} - \left(\tilde{x}^{(0)}(k)\right)^{1/\alpha} \right|. \quad (3.7)$$

With the principle of the minimum cumulative error of the fitted values and actual values, combined with the application of trigonometric wave mode matching method, we try to make sure that the posteriori error C is the smallest and the small probability P is the largest, and then we obtain $\omega = 1.06$. At the same time, the amplitude of wave mode calculated by the formula (3.7) is $A = 3.14$.

Take A and ω into (3.6); then we obtain the revised residuals' formula:

$$f = 3.14 \cdot \sin(1.06 \cdot (t-1)). \quad (3.8)$$

Combined with residual formula and the formula (3.5), the final forecast expression after residuals adjustment is

$$x^{(0)}(k) = \left(x^{(1)}(k) - x^{(1)}(k-1)\right)^{1/0.2} - 15 + 3.14 \cdot \sin(1.06 \cdot (t-1)). \quad (3.9)$$

Let us predict track cross level state with the formula (3.9). The predicted values and actual values are shown in Figure 7.

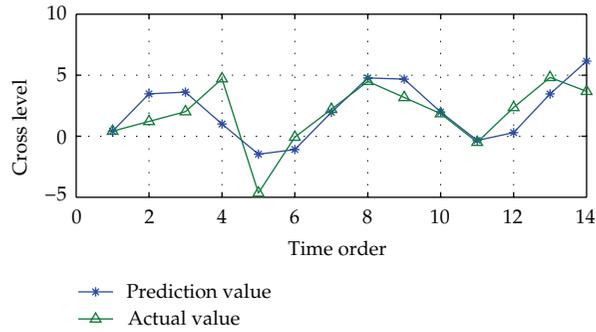


Figure 7: Comparison of actual value and revised predictive value at fixed measuring point.

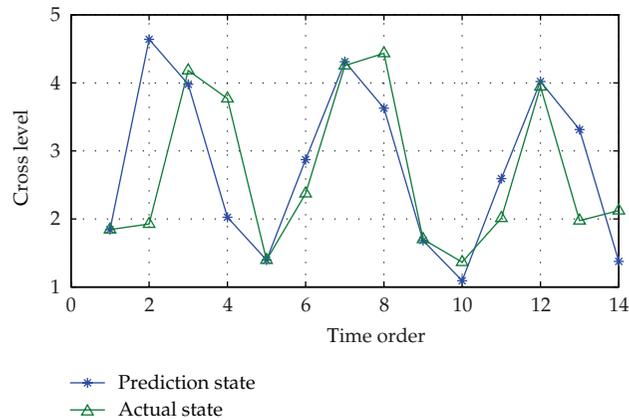


Figure 8: Comparison of actual state and prediction state at unit section after residual modification.

When gray model GM (1,1) after residual modification is used to forecast time series data (see Section 2.2) at the unit section, the prediction formula is (3.10), and the prediction result is shown in Figure 8:

$$x^{(0)}(k) = \left(x^{(1)}(k) - x^{(1)}(k-1) \right)^{1/0.2} + 1.6 \cdot \sin(1.26 \cdot (t-1)). \quad (3.10)$$

As can be seen from Figures 7 and 8, compared to the original forecasting trend curve, the modified forecasting trend curve is much closer to the actual value. It has a better degree of fitting and can reflect the cyclical changes of the track cross level state. Therefore, the revised model can be applied to forecast the future track cross level state trends in the medium and in the long term.

In gray forecasting, the prediction with good fitting and extrapolation leads to a smaller value C and a larger value P . It shows a large probability of small error and high accuracy in prediction [20]. According to the statistical theory, we examine the accuracy of prediction on track state by using posteriori error C and the small probability P and then make a comparison between it and the predictive accuracy of original GM (1,1) model. See Table 2.

Table 2: Comparison of model's accuracy.

Test items	GM (1, 1)	GM (1, 1) after residual modification
C	65%	43%
P	86%	86%

Table 3: Values of autocorrelation function, autocovariance function, and partial incidence degree function.

k	0	1	2	3	4	5	6	7
$\tilde{\rho}_k$	1	-0.5910	0.0835	0.0075	0.0040	-0.1106	0.3729	-0.4402
\tilde{r}_k	20.7675	-12.2731	1.7341	0.1567	0.0821	-2.2967	7.7441	-9.1408
$\tilde{\varphi}_{kk}$	1	-1.0356	-0.9201	-0.7259	-0.5022	-0.2565	0.1265	-0.0529

Through comparative analysis, the variance ratio of posteriori error of GM (1, 1) model after the residual modification is significantly smaller than the original residual model; thus the fitting and extrapolation of the modified model have changed for the better, and the predictive accuracy is improved.

4. Short-Term Prediction Models of Cross Level State Change

4.1. Prediction Based on AR Model

Track cross level irregularity time series data is smooth and consistent with the characteristics of the stationary random sequence; so there is no need to eliminate the trend of the differential operator. Although there is no definite model in track state changes in the long run, the state change in a short period can still be considered as close to the linear model. In order to study the unit section of the overall level of state which changes over time, it is considered as one-dimensional array data which contains 300 data at a select unit section. The track cross level irregularity time series data is

$$Z_t = [z_t^0, z_t^1, \dots, z_t^m], \quad t = 1, 2, \dots, 31, \quad m = 1, 2, \dots, 300. \quad (4.1)$$

Then, $\omega_t = Z_t - \bar{Z}$, time series $\{\omega_t\}$ is generated with mean value zero, $t = 1, 2, \dots, 31$, and \bar{Z} is the sample mean value of Z_t .

Autocovariance function refers to the random signal between the values of two different moments of the second-order mixed central moments. Autocorrelation function depicts the incidence degree between adjacent variables of time sequence. The partial autocorrelation function was excluded from the impact of other intermediate variables; the two functions are closely related and can reflect the true incidence degree between two variables [21]. The value of sample's autocovariance function \tilde{r}_k , incidence degree function $\tilde{\rho}_k$, and partial autocorrelation function $\tilde{\varphi}_{kk}$ are shown in Table 3.

It can be seen from Table 3 that when k becomes greater, the previous four $|\tilde{\rho}_k|$, absolute value of $\tilde{\rho}_k$, are getting smaller and smaller. Therefore, we can see that the autocorrelation function $\tilde{\rho}_k$ is tailed. When $k > 4$, there is at most one $\tilde{\varphi}_{kk}$ that can make $|\tilde{\varphi}_{kk}| \geq 0.3591$. Therefore, the sample partial incidence degree function $\tilde{\varphi}_{kk}$ is truncated; partial autocorrelation function $\tilde{\varphi}_{kk}$ is truncated at the point at which $k = p = 4$.

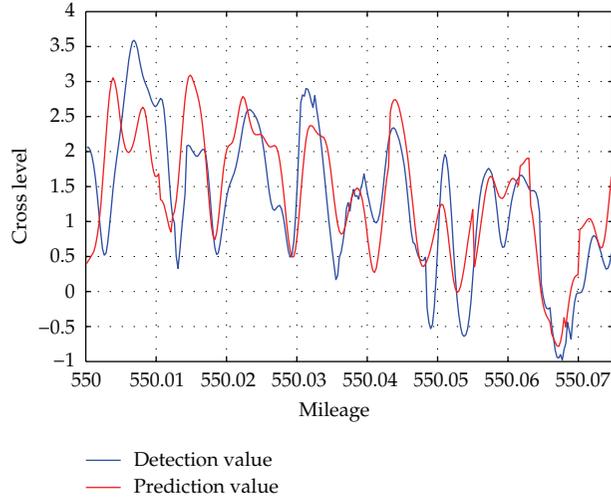


Figure 9: Comparison of predicted value and actual inspection data in late June 2009.

Through comprehensive analysis, the prediction model is defined as AR (4). We can calculate the parameter's estimates $\tilde{\varphi}$ using Yule-Walker equations.

We get $\tilde{\varphi}_1 = -1.0257$, $\tilde{\varphi}_2 = -0.8123$, $\tilde{\varphi}_3 = -0.5224$, and $\tilde{\varphi}_4 = -0.2292$.

Thus, the AR (4) model is

$$Z_t - \bar{Z} = \tilde{\varphi}_1 (Z_{t-1} - \bar{Z}) + \tilde{\varphi}_2 (Z_{t-2} - \bar{Z}) + \tilde{\varphi}_3 (Z_{t-3} - \bar{Z}) + \tilde{\varphi}_4 (Z_{t-4} - \bar{Z}) + a_t. \quad (4.2)$$

In the formula, $t = k + l$, a_t is the random disturbance error, which is white noise sequences with zero mean value, normally variance, nonzero, unrelated, and independent.

Taking estimated value on both sides of formula (3.8) and then take the estimated parameter into the formula, we can get the AR (4) prediction formula:

$$\tilde{Z}_{k+l} = -1.0257\tilde{Z}_{k+l-1} - 0.8123\tilde{Z}_{k+l-2} - 0.5224\tilde{Z}_{k+l-3} - 0.2292\tilde{Z}_{k+l-4} - 1.5896\bar{Z}. \quad (4.3)$$

In formula, when $k + l - i < n$, $\tilde{Z}_{k+l-i} = Z_{k+l-i}$.

The predictive results of cross level irregularity data at late June 2009 and actual test data are shown in Figure 9.

By contrasting the forecasted data with the actual inspection data, it can be found that the distribution characteristics of actual value and the predictive value can agree with each other well, and the data curves roughly coincide with each other.

4.2. Prediction Model Based on Kalman Filtering

Kalman filtering can be used to estimate the current state when the estimated state from the last time and the current state are known, needless to know historical information observations or estimates. In the absence of maintenance, changes of track geometry are closely related to the passing gross weight change; deviation of track geometry will be further

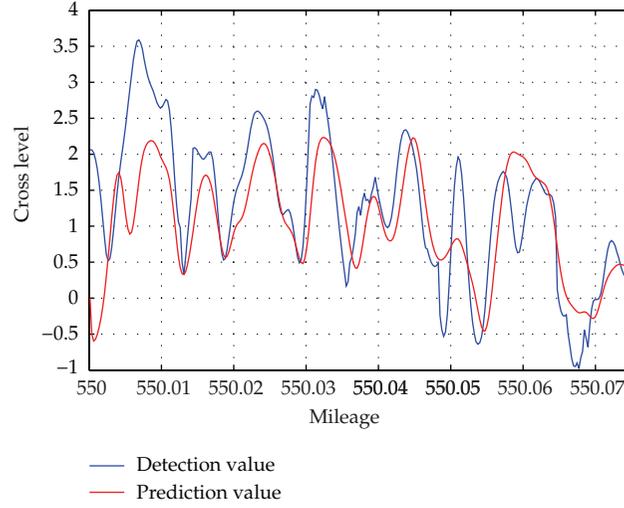


Figure 10: Comparison of detection value and the prediction value (ANN).

from the standard value with the increase in gross weight; track geometry status is also affected by the impact of train speed. The higher the speed is, the greater force is exerted on the track and the greater influences on the track geometry status are. The track geometry (detection data), passing gross weight change, and train speed are used as the technical indicators for track state prediction, and the accumulation and analysis of historical data can be used for building track state prediction models.

With the application of Kalman filtering algorithm, in the track inspection data analysis and forecasting models, x_k is actual value of track inspect items; A is the transfer matrix of the actual value; z_k is track inspection car's detecting value; H is the observation matrix; w_k is process noise $p(w) \sim N(0, Q)$, which is the deviation of the track state changes; v_k is measuring noise $p(v) \sim N(0, R)$, which is white Gaussian noise; $\hat{s}(k)$ is the predictive value of the state of the track geometry. The prediction formula of Kalman filtering is as follows:

$$\hat{s}(k) = A\hat{s}(k-1) + H(k)[X(k) - C(k)A(k)\hat{s}(k-1)]. \quad (4.4)$$

In the formula, $\hat{s}(k)$ is prediction value, $H(k)$ is the minimum mean-square deviation under the revised matrix, $A(k)$ is transfer matrix, $X(k)$ is an observed value and $\hat{s}(k)$ is an estimated value, and $C(k)$ is the measured matrix.

Kalman filter model is applied to forecast the cross level status the next time when testing. The comparison of detection cross level value and the prediction value is shown in Figure 10.

4.3. Prediction Model Based on Artificial Neural Network

Artificial neural network (ANN) is widely used in function approximation, pattern recognition, and data compression [22–24]. It is the best method compared with other traditional models, because it has better durability, timely forecasts, highly nonlinear, and

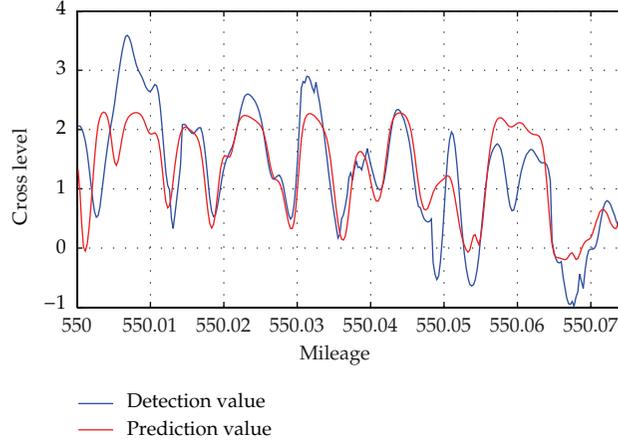


Figure 11: Comparison of the detection value and the prediction value (Kalman filter model).

strong self-adaptive learning ability. Usually, the network has an input layer, an output layer, and a hidden layer. ANN has advantages such as the following. Network's input and output can be achieved in any nonlinear mapping as long as there are enough hidden layers and hidden nodes. The relationship between input nodes and output nodes of ANN is as follows [10]:

$$y_t = w_0 + \sum_{j=1}^Q w_j \cdot g \left[\sum_{i=1}^p w_{ij} \cdot y_{t-i} + w_{0j} \right] + e_t. \quad (4.5)$$

In the formula, w_{ij} ($i = 1, 2, \dots, p$, $j = 1, 2, \dots, Q$), $\text{sig}(x) = (1/(1 + e^{-x}))$ is a hidden layer transfer function, y_t is actual output, w_{ij} is connection weights, p is the number of input nodes, and Q is the number of hidden nodes. The neural network model performs a nonlinear functional mapping from the past observations to the future value y_t , that is,

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, W) + e_t. \quad (4.6)$$

W is a vector of all parameters and $f(\cdot)$ is a function determined by the network structure and connection weights. Because track state changes are nonlinear and ANN has a flexible capability in nonlinear modeling, ANN is applied to forecast the track data change. The forecasting result is shown in Figure 11.

4.4. Comparison of Three Prediction Models

The specific error distribution of AR model, Kalman filtering model, is ANN model are shown in Table 4.

It can be seen from Table 4 that the predictive accuracy of AR and Kalman filtering models is similar, and the predictive accuracy of ANN model is slightly higher than the previous two.

Table 4: Error distribution of forecasted data.

Models	Error range	$ r > 1$	$0.5 < r \leq 1$	$0.2 < r \leq 0.5$	$0 \leq r \leq 0.2$
AR	Amount	44	104	82	70
	Percentage	15%	35%	27%	23%
Kalman filtering	Amount	37	99	87	77
	Percentage	12%	33%	29%	26%
ANN	Amount	36	65	99	100
	Percentage	12%	22%	33%	33%

5. Conclusions

After the comprehensive assessment of the incidence degrees of track irregularity between various indicators of factors, we find that when the associated values are higher, these correlated time sequences will normally have a higher degree of factors correlation or processes correlation. Meanwhile, the calculated results of incidence degree will be in a good agreement with the actual situation, which will provide a reliable basis for choosing modeling variables and analyzing factors. Improved GM (1, 1) model based on features of track cross level data can predict track state development and changes at fixed measuring point in the medium and long term. Fitting curve can reflect the cyclical changes of cross level state over time by residual modification. Statistical validation shows that the posteriori error values in improved model which was corrected with residuals will be reduced down from 65% to 43%, compared to the original model. It reflects the changes of cross level state more accurately. Random linear AR model, Kalman filtering, and ANN are used to predict the state changes of union section in short term. The results show that the accuracy of ANN is slightly higher than AR model and Kalman filtering, and the combination of the four models together constitutes the research of long-term and short-term track state changes at fixed measuring point and union section.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (General Projects) (Grant no.: 61272029), National Key Technology R&D Program (Grant no.: 2009BAG12A10), China Railway Ministry Major Program (2008G017-A), and State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University (Contract no.: RCS2009ZT007).

References

- [1] D. Julong, *The Primary Methods of Grey System Theory*, Huazhong University of Science and Technology Press, Wuhan, China, 2005.
- [2] G. Liu and J. Yu, "Gray correlation analysis and prediction models of living refuse generation in Shanghai city," *Waste Management*, vol. 27, no. 3, pp. 345–351, 2007.
- [3] M. Marcellino, J. H. Stock, and M. W. Watson, "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," *Journal of Econometrics*, vol. 135, no. 1-2, pp. 499–526, 2006.
- [4] J. Ding, L. Han, and X. Chen, "Time series AR modeling with missing observations based on the polynomial transformation," *Mathematical and Computer Modelling*, vol. 51, no. 5-6, pp. 527–536, 2010.

- [5] R. Kalman, "A new approach to linear filtering and prediction problems," *Transaction of the ASME-Journal of Basic Engineering Series D*, vol. 82, pp. 35–45, 1960.
- [6] B. Feil, J. Abonyi, S. Nemeth, and P. Arva, "Monitoring process transitions by Kalman filtering and time-series segmentation," *Computers and Chemical Engineering*, vol. 29, no. 6, pp. 1423–1431, 2005.
- [7] R. Kandepu, B. Foss, and L. Imsland, "Applying the unscented Kalman filter for nonlinear state estimation," *Journal of Process Control*, vol. 18, no. 7-8, pp. 753–768, 2008.
- [8] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing, Explorations in the Microstructure of Cognition*, MIT Press, 1986.
- [9] P. P. Balestrassi, E. Popova, A. P. Paiva, and J. W. Marangon Lima, "Design of experiments on neural network's training for nonlinear time series forecasting," *Neurocomputing*, vol. 72, no. 4–6, pp. 1160–1178, 2009.
- [10] M. Khashei, M. Bijari, and G. A. Raissi Ardali, "Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs)," *Neurocomputing*, vol. 72, no. 4–6, pp. 956–967, 2009.
- [11] P. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- [12] H. Liu, H.-Q. Tian, and Y.-F. Li, "Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction," *Applied Energy*, vol. 98, pp. 415–424, 2012.
- [13] P. Areekul, T. Senjyu, H. Toyama, and A. Yona, "A hybrid ARIMA and neural network model for short-term price forecasting in deregulated market," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 524–530, 2010.
- [14] M. Khashei, S. R. Hejazi, and M. Bijari, "A new hybrid artificial neural networks and fuzzy regression model for time series forecasting," *Fuzzy Sets and Systems*, vol. 159, no. 7, pp. 769–786, 2008.
- [15] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Systems with Applications*, vol. 37, no. 1, pp. 479–489, 2010.
- [16] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," *Applied Mathematics Letters*, vol. 22, no. 9, pp. 1467–1470, 2009.
- [17] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research Part C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [18] W. Wang, *Vehicle's Man-Machine Interaction Safety and Driver Assistance*, China Communications Press, Beijing, China, 2012.
- [19] X. Xinpeng, "Theoretical study and reviews on the computation method of grey interconnect degree," *Systems Engineering-Theory & Practice*, vol. 17, no. 8, pp. 77–82, 1997.
- [20] J. H. Stock and M. W. Watson, "Implications of dynamic factor models for VAR analysis," NBER Working Paper no. 11467, 2005.
- [21] N. L. Yu, D. Y. Yi, and X. Q. Tu, "Analyzing auto-correlation and partial-correlation functions in time series," *Mathematical Theory and Applications*, vol. 27, no. 1, pp. 54–57, 2007.
- [22] M. Khayet, C. Cojocar, and M. Essalhi, "Artificial neural network modeling and response surface methodology of desalination by reverse osmosis," *Journal of Membrane Science*, vol. 368, no. 1-2, pp. 202–214, 2011.
- [23] A. Norets, "Estimation of dynamic discrete choice models using artificial neural network approximations," *Econometric Reviews*, vol. 31, no. 1, pp. 84–106, 2012.
- [24] C. B. Cai, H. W. Yang, B. Wang, Y. Y. Tao, M. Q. Wen, and L. Xu, "Using near-infrared process analysis to study gas-solid adsorption process as well as its data treatment based on artificial neural network and partial least squares," *Vibrational Spectroscopy*, vol. 56, no. 2, pp. 202–209, 2011.

Research Article

Research on Analysis Method of Traffic Congestion Mechanism Based on Improved Cell Transmission Model

Hongzhao Dong, Shuai Ma, Mingfei Guo, and Dongxu Liu

The Joint Institute of Intelligent Transportation System, Zhejiang University of Technology, Hangzhou 310014, China

Correspondence should be addressed to Hongzhao Dong, its@zjut.edu.cn

Received 8 August 2012; Accepted 21 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Hongzhao Dong 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 analyze the spreading regularity of the initial traffic congestion, the improved cell transmission model (CTM) is proposed to describe the evolution mechanism of traffic congestion in regional road grid. Ordinary cells and oriented cells are applied to render the crowd roads and their adjacent roads. Therefore the traffic flow could be simulated by these cells. Resorting to the proposed model, the duration of the initial traffic congestion could be predicted and the subsequent secondary congestion could be located. Accordingly, the spatial diffusion of traffic congestion could be estimated. At last, taking a road network region of Hangzhou city as an example, the simulation experiment is implemented to verify the proposed method by PARAMICS software. The result shows that the method could predict the duration of the initial congestion and estimate its spatial diffusion accurately.

1. Introduction

The urban traffic congestion has been becoming a more and more serious issue, especially in China. Traffic congestion is a typical traffic condition with a dynamic course by time. The congestion generating mechanisms analysis considering both time and space scale helps to illustrate the dynamic changing characteristic of the traffic flow that eventually leads to the congestion. Many scholars had studied the urban traffic congestion by cell transmission model (CTM) at the 1990s [1–3]. Some scholars, such as Szeto, Jia Bin et al., had researched CTM model to have a good performance in simulating the typical dynamic characteristics of traffic flow, such as the formation of shock waves, traffic congestion, and the dynamic evolution rule of node combined by Multiroad [4, 5]. Certainly, other scholars, such as Chen Qian, ZHOU Xi-peng, and YANG Zhao-shen, applied CTM model to the models of traffic

flow spread on network such as traffic bottleneck recognition and modeling for traffic events duration time and traffic recovery time [6–10].

In previous studies, the spatial diffusion estimating and duration predicting of traffic congestion are separately supposed as both independent issues on congestion. However, both issues usually affect and restrain one another. Actually they have remarkable relationship in time and space scale. Therefore, the analysis method based on our proposed improved CTM model will give a new solution to track traffic congestion considering both time and space scale. And the research in this paper promises to prevent and relieve traffic congestion and improve the utilization efficiency of the road resources.

This paper is organized as follows. The spatial diffusion regularity of traffic congestion is analyzed in Section 2. The improved cell transmission model is given to describe the spread of traffic flow in the road grid in Section 3. The analysis method of traffic congestion mechanism based on improved CTM is researched in Section 4. In Section 5, taking a road network region of Hangzhou city as an example, the experimental simulation is provided to demonstrate the application of our proposed method by PARAMICS software. The last section highlights the conclusion.

2. Spatial Diffusion Analysis of Traffic Congestion

Traffic congestion refers to traffic flow detained phenomenon caused by the contradiction between traffic requirement and traffic capacity. Generally, traffic congestion would occur if the traffic capacity supplied by traffic facilities is close to or less than the current traffic demand. In some specific time, the regular bottlenecks will be formed if the traffic demand keeps being beyond the transportation capacity. Such case is called regular traffic congestion. Besides, if the actual capacity declines because of temporary event or accident, occasional traffic congestion is generated.

For a jammed unit in the road network, if the traffic demands remain high level and even continue to increase, the actual traffic capacity to collect the upstream traffic flow will become more and more limited. It leads to slow traffic speed and quickly rising traffic density of the upstream unit. That is how the primary traffic congestion can spread to upstream unit. In addition, traffic flow is usually controlled by traffic signal control system. Consequently, traffic flow in jammed unit with small headway will spread to the downstream unit during the green time. Traffic demand and v/c of downstream unit will rise rapidly. With the increasing interference between the vehicles with one another, the traffic speed will be dropped and it may also lead to the possible congestion. These factors easily make the traffic flow unstable and the capacity to drop. It is why the downstream units may be jammed. Therefore, the original congestion can spread to the whole road grid quickly. If there are no effective corresponding measures to relieve its spatial diffusion, the whole or regional traffic congestion will occur.

3. Improved Cell Transmission Model

3.1. Cell Transmission Model

Cell transmission model (CTM) is based on assumption that the traffic flow q and density ρ have a relationship as shown in Figure 1.

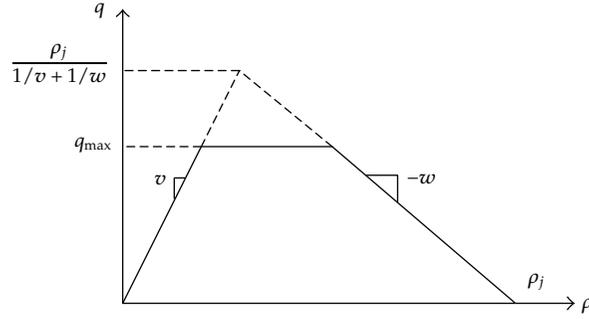


Figure 1: A trapezoidal fundamental diagram for the cell transmission model.

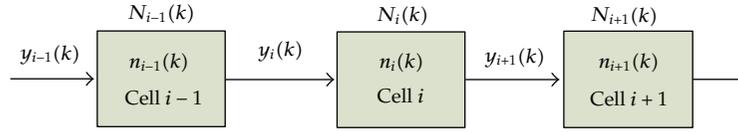


Figure 2: Traffic flow moving model of CTM.

CTM means the road section could be divided into several cells and time series discretized into several time steps. The length of each cell is the driving distance with free traffic flow during one time step. It should satisfy the below formula to describe the traffic flow between two nearby cells by CTM:

$$q = \min\{v\rho, q_{\max}, w(\rho_j - \rho)\}. \quad (3.1)$$

Here v is free traffic flow speed and w is back propagation speed of traffic wave when it is jammed. q_{\max} is maximum traffic volume. Under low traffic density, v is constant. While traffic density is high, w is constant and $v > w$. For $0 < \rho < \rho_j$, the above quantity is satisfied.

The traffic flow spreading process between cells is shown as Figure 2. At time step k , $y_i(k)$ is the total flow entering the cell i . $N_i(k)$ is the maximum capacity of the cell i . $n_i(k)$ is the number of vehicles in the cell i .

$q_i(k)$ is the inflow rate of the cell i and $q_{i-\max}(k)$ is the maximum inflow rate of the cell i . And the function relation between vehicles and the density of cell is as the following equation:

$$n_i(k) = \rho_i(k)v\delta. \quad (3.2)$$

So it is deduced the traffic flow spread relationship between two adjacent cells is as (3.3). $Q_i(k)$ is the maximum inflow rate of cell i . Then after discretizing it, traffic flow conservation equation in CTM is as (3.4):

$$y_i(k) = \min\left\{n_{i-1}(k), Q_i(k), \frac{\omega(N_i(k) - n_i(k))}{v}\right\}, \quad (3.3)$$

$$n_i(t+1) = n_i(t) + y_i(t) - y_{i+1}(t). \quad (3.4)$$

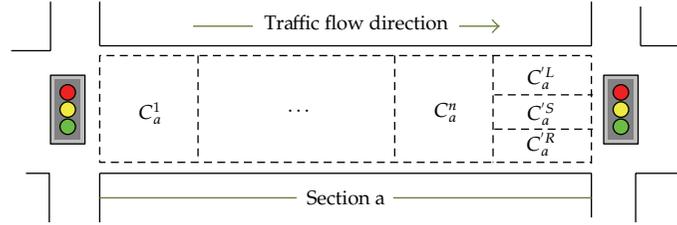


Figure 3: Cell definition improved CTM.

The existing CTM method is appropriated for describing traffic flow of linear road such as freeway. Unfortunately, the weakness of node diversion and node confluence appears in the existing CTM and it is unable to suit for traffic flow modeling of urban road network. Accordingly, the improved CTM method is researched as below.

3.2. Improved CTM and Its Cell Definition

The spread of traffic flow in the road grid could be classed into two aspects. One is traffic flow transmission inside the road section and another is between the adjacent sections. Recently many scholars use Daganzo's cell transmission model to construct the spread of traffic flow which includes section model and node model. The node model involves the signal control and oriented lanes at the end of the section. Based on the node model, the improved section model is proposed with oriented cell instead of node model into section model to realize the spread of traffic flow. It helps to simplify the modeling process and improve the calculation efficiency.

Oriented cell is defined as the cell locating at oriented lane of section. It plays the role of traffic flow diverging. The capacity and outflow of oriented cells is decided by the timing scheme of the downstream traffic control system. Therefore the key parameters of oriented cell are different because of their different traffic flow direction.

According to the principle of lane group classification, define C_k^L , C_k^S , and C_k^R as three kinds of oriented cells to describe spatial distribution of traffic flow shown as Figure 3. It is essential to analyze the spatial diffusion of traffic congestion.

3.3. Parameter Definition and Constraint Condition

Along the traffic direction, section a consists of ordinary cells C_a^i ($i = 1, 2, \dots, m$; i is numbered in ascending order from 1 along the traffic direction.) and a series of oriented cells. At time step k , parameter of two types of cell is shown as in Table 1.

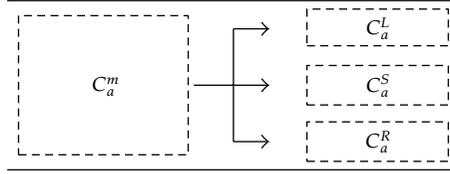
According to the core theory of CTM, traffic flow transmission between cells could be classified in the three types below.

The first type is the flow expression of ordinary cells in the same section shown as in (3.3).

The second type is the flow expression of ordinary cell and oriented cells in the same section shown in (3.5), (3.6), and (3.7). This type describes that the traffic flow

Table 1: Parameter definition of model.

Parameter	Ordinary cell	Oriented cell
Density	$\rho_a^i(k)$	$\rho_a^\xi(k)$
Vehicles	$n_a^i(k)$	$n_a^\xi(k)$
Inflow rate	$q_a^i(k)$	$q_a^\xi(k)$
Inflow	$y_a^i(k)$	$y_a^\xi(k)$
Max. outflow	$q_a^{i,\max}(k)$	$q_a^{\xi,\max}(k)$
Max. inflow	$Q_a^i(k)$	$Q_a^\xi(k)$
Capacity	$N_a^i(k)$	$N_a^\xi(k)$

**Figure 4:** Flow transmission between ordinary cell and oriented cells.

of the ordinary cell moves into different oriented cells on the basis of trip demand shown in Figure 4:

$$y_a^L(k) = \min \left\{ n_a^m(k) \eta_L, Q_a^L(k), \frac{\omega [N_a^L(k) - n_a^L(k)]}{v} \right\}, \quad (3.5)$$

$$y_a^R(k) = \min \left\{ n_a^m(k) \eta_R, Q_a^R(k), \frac{\omega [N_a^R(k) - n_a^R(k)]}{v} \right\}, \quad (3.6)$$

$$y_a^S(k) = \min \left\{ n_m(k) (1 - \eta_L - \eta_R), Q_a^S(k), \frac{\omega [N_a^S(k) - n_a^S(k)]}{v} \right\}, \quad (3.7)$$

η_L, η_R are the behavior proportion of left turn and right turn of traffic flow in the section.

The third type is the flow expression of oriented cell and ordinary cells of nearby sections shown in (3.8). This type describes that the traffic flow of the oriented cell moves into ordinary cells of different downstream sections shown in Figure 5:

$$Q_a^\xi(k) = \min \left\{ \frac{g_a^\xi(k)}{h(v)}, \frac{\omega [N_a^\xi(k) - n_a^\xi(k)]}{v}, q_a^{\xi,\max} \delta \right\}, \quad (3.8)$$

ξ can express $L, S,$ or R . $g_a^\xi(k)$ is green time in the direction of traffic flow in oriented cell by signal control system. $h(v)$ is the average headway of free traffic flow.

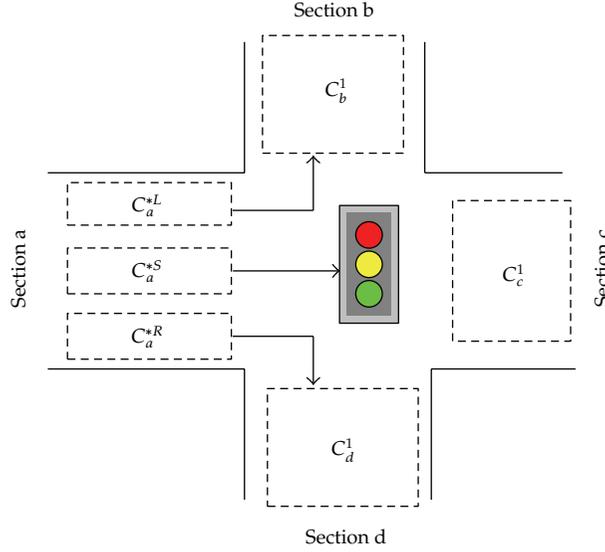


Figure 5: Flow transmission between oriented cells and ordinary cells.

Therefore flow expression is updated in (3.9):

$$y_b^1(k) = \min \left\{ n_b^1(k), Q_a^{\xi}(k) \right\}. \quad (3.9)$$

If the model is applied to urban road grid, the oriented cell and the ordinary cell is difficult to be defined as equal length. It will lead to the strong noise of model. The traffic density conservation function is introduced to solve the issue. The function can ignore the cell length inequality that leads to disorder of traffic flow transmission. The general formula is shown in (3.10):

$$\rho_i(k+1) = \rho_i(k) + \frac{\delta}{l_i} (y_i(k) - y_{i+1}(k)). \quad (3.10)$$

Traffic density conservation formula for the second and third type is shown in (3.11):

$$\frac{\sum_{\xi=R,S,L} \rho_a^{\xi}(k+1)}{3} = \frac{\sum_{\xi=R,S,L} \rho_a^{\xi}(k)}{3} + \frac{\delta}{l_a^{\xi}} [y_a^m(k) - y_b^1(k)]. \quad (3.11)$$

4. Analysis Method for Traffic Congestion Mechanism Based on Improved CTM

The analysis method for traffic congestion mechanism contains both of predicting and locating, that is, to predict the duration of initial congestion and to locate secondary congestion. To improve the accuracy of the method, the sections can be divided into three groups according to their locations and the traffic flow direction. The first one is the object group that is being

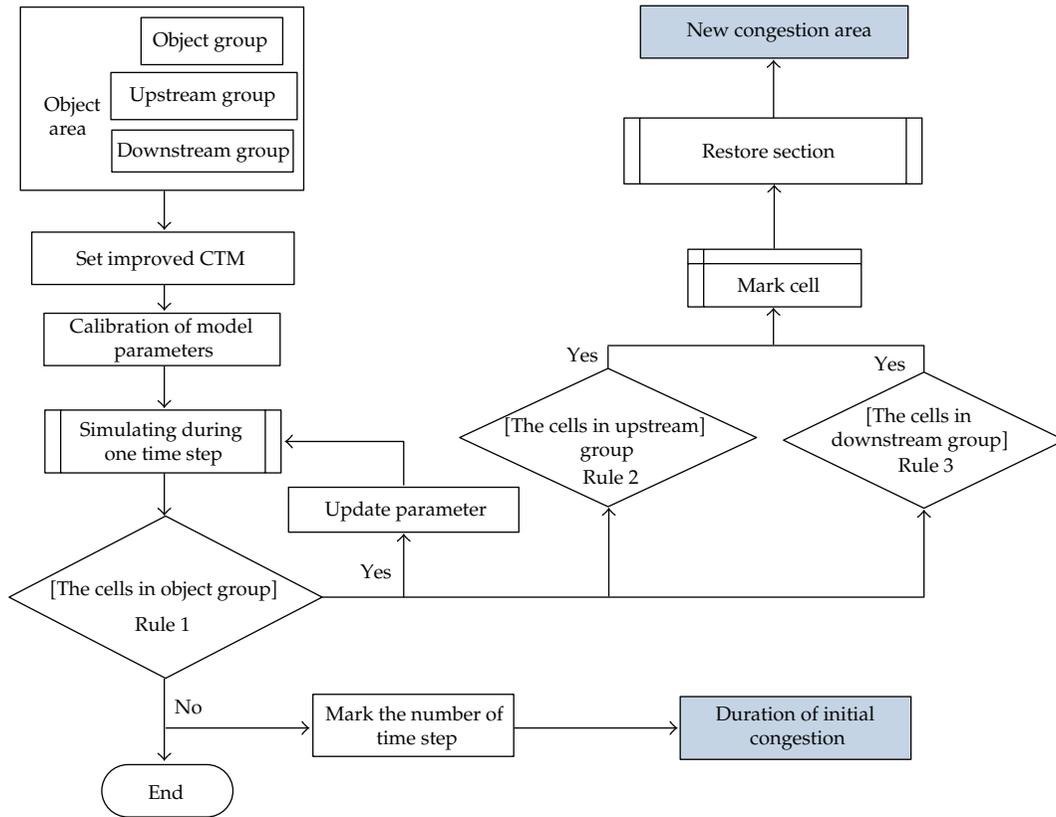


Figure 6: Flowchart of analysis method for traffic congestion mechanism based on CTM.

jammed sections. The second group is the upstream group including the sections locating at upstream of the object group. The third group is the downstream group that locates downstream of object section.

The analysis method for traffic congestion mechanism based on CTM is as follows. The analytical cycle is discretized into several time steps. The dynamic traffic characteristic is simulated by the inflow vehicles and outflow vehicles during the period of one time step. After every time step is completed, it will repeatedly output the value of the key variables such as the inflow vehicle quantity and the traffic capacity of the section until the termination conditions are met. After each step, the cells in different groups are judged by the corresponding rules. If the cell complies with the rules, it will be marked until the end condition. The marked cells will restore the section. The outcome of new congestion area and duration of initial congestion can be obtained eventually. The flow chart of the method is shown in Figure 6.

All cells in object group are analyzed to predict the duration of the initial congestion. If all cells appear to be noncongestion status after λ time steps, the duration of initial congestion is as (4.1):

$$T_{\text{last}} = \lambda\delta. \tag{4.1}$$

The ordinary cells in the second group and the third group are analyzed to locate the secondary congestion. After repeatedly simulating and comparing with the practice data, the judgment rules are introduced to indentify traffic state of cell. Rule 1 *describes* the judgment rule of the cell of section p in the object group.

Rule 1. If there is $y_p^{i+1}(k) = n_p^i(k)$, then $C_p^i(k) \notin U_{cgt}(k)$. Otherwise $C_p^i(k) \in U_{cgt}(k)$.

$U_{cgt}(k)$ is the set of congestion cells after No. k time step. When $U_{cgt}(\lambda)$ is empty, it means the initial congestion ends.

Rules 2 and 3 describe the judgment rules of the cell of section a in the upstream group and the cell of section b in the upstream group during the period T_γ which is the duration of initial congestion.

Rule 2. If there is $f(T_\gamma) = \text{count}[y_a^i(k) < n_a^i(k)] \geq \varepsilon_{\text{exp}1}$, then $C_a^i(k) \in U_{cgt}(T_\gamma)$. Otherwise $C_a^i(k) \notin U_{cgt}(T_\gamma)$.

Rule 3. If there is $f(T_\gamma) = \text{count}\{y_b^i(k) = [N_b^i(k) - n_b^i(k)]\} \geq \varepsilon_{\text{exp}2}$, then $C_b^i(k) \in U_{cgt}(T_\gamma)$. Otherwise $C_b^i(k) \notin U_{cgt}(T_\gamma)$.

$f(T_\gamma) = \text{count}(\text{expression})$ means the number of variables of cell satisfy the *expression* after each time step. ε_{exp} is the default threshold.

Above all, the analysis method for traffic mechanism based on CTM can be divided into four steps.

Step 1. Define the cells of three groups and assign values to basic variables.

Step 2. Initialize the numbers of vehicles, the capacity, and other variables of each cell.

Step 3. The vehicles transmit between cells referring to (3.5), (3.6), (3.7), and (3.9). Cells are divided into sets according to Rules 1–3 after each time step.

Step 4. Repeat Step 3 until satisfying the terminating condition.

5. Simulation Experiment of Our Proposed Method by PARAMICS

PARAMICS is the reliable, feature-rich microscopic traffic simulation software, which is widely applied to research intelligent transportation. The simulation experiment contains road network modeling supported by modeller module and API programming supported by programmer module.

The road network model has been constructed referring to the location and spatial relationship of real road network in Hangzhou. The traffic control system of each node in network is set by the typical actual signal timing plan. The traffic detector is fixed at the beginning of each cell to virtually realize the cell definition in the model shown in Figure 7. The traffic flow acquired by the detector can promise to verify the validity and feasibility of the proposed method.

Here are some parameters in the experiment. The congestion density of each lane is 142 puc/km. The free-flow speed is 36 km/h. Time step is 10 s. The length of ordinary cell is 100 m. The capacity of each lane in ordinary cell is 14 puc. Assuming at the beginning the section of Tiyuchang Road from Wulin Crossing to Yan'an Crossing is jammed and the traffic congestion is triggered by short-term traffic demand expansion.

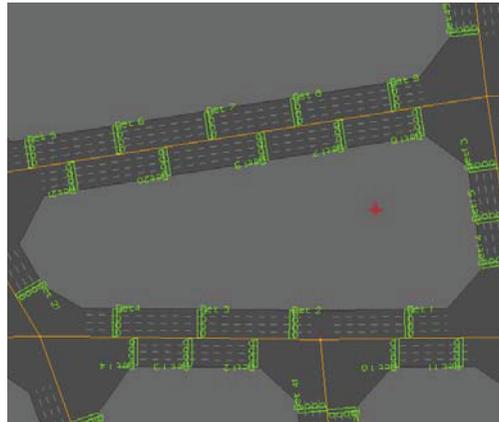


Figure 7: Virtual cell of the network.

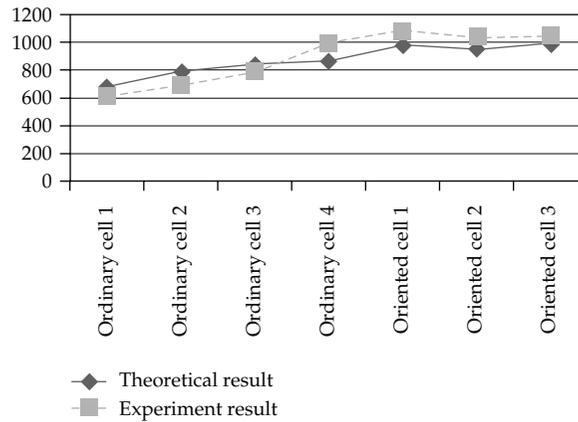


Figure 8: Comparing experiment and theoretical results.

The experiment results of duration of initial congestion are compared with the theoretical value shown in Figure 8.

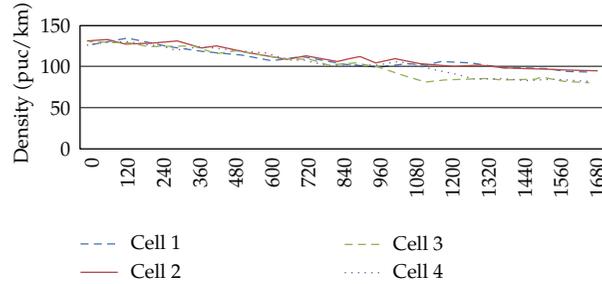
The density of each ordinary cell in object group is shown in Figure 9. The above curves shows that cell1, cell2, cell3, and cell4 separately have obvious downward trend during [600 s–720 s], [720 s–840 s], [840 s–960 s], [960 s–1080 s]. Both of the duration of initial congestion in the experiment and its theoretical result locates in these areas.

The above results demonstrate that the max duration of congestion is 930 s of all cells in the object group. Therefore during 96 time steps, the traffic state of sections of upstream and downstream groups is influenced by initial congestion shown in Table 2.

The above table suggests that the influence to upstream section by initial congestion is more than one to downstream section. Vehicles queuing phenomenon is the most significant characteristic of congestion. The vehicles from upstream join the queue and the queue is lengthened into upstream section. Therefore, the congestion spreads to upstream sections. Besides, the vehicle queue is intermittently starting and stopping because of traffic control system. The influence to downstream section is related to the signal time plan by traffic control system. What is more, the improved CTM have good performance on modeling traffic

Table 2: Results of second congestion location.

Time step	The number of congestion cells in upstream group		The number of congestion cells in downstream group		Precision	
	Theoretical result	Experiment result	Theoretical result	Experiment result	Matching number	Precision rate
1–12	3	3	1	0	2	91.67%
13–24	5	5	2	2	4	83.33%
25–36	6	7	2	3	6	83.33%
37–48	6	6	3	3	7	87.5%
49–60	5	6	3	3	6	83.33%
61–72	5	4	2	2	5	91.67%
73–84	4	3	2	1	4	91.67%
85–96	3	2	1	1	2	87.5%

**Figure 9:** Density curves of ordinary cells in object group.

flow of the urban road grid and describing the congestion issue considering time and space scales at the same time comparing the CTM.

6. Conclusion

The improved cell transmission model with the oriented cell instead of node model is proposed to realize the spreading of traffic flow in the road grid, which could not only simplify the model but also improve the adaptability of the model. The analysis method of traffic congestion mechanism based on the model could be applied to predict the duration of the initial congestion and locate the secondary congestion. Besides, the micro-simulation experiments demonstrate the validity and feasibility of our proposed comprehensive method, which can satisfy the analytical requirements of traffic congestion in the urban transportation.

Acknowledgment

The research is support by National Natural Science Foundation of China (Grant no. 61174176).

References

- [1] C. F. Daganzo, "The lagged cell transmission model," in *Transportation and Traffic Theory*, A. Ceder, Ed., pp. 81–103, Pergamon-Elsevier, New York, NY, USA, 1999.
- [2] 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.
- [3] W. Y. Szeto, "Enhanced lagged cell-transmission model for dynamic traffic assignment," *Transportation Research Record*, no. 2085, pp. 76–85, 2008.
- [4] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez, "Piecewise-linearized cell transmission model and parameter calibration methodology," *Transportation Research Record*, no. 1965, pp. 183–191, 2006.
- [5] B. Jia, Z. Y. Gao, K. P. Li, and X. G. Li, *Models and Simulations of Traffic System Based on Theory of Cellular Automaton*, Science Press, Beijing, China, 2007.
- [6] Z. S. Yang, X. Y. Gao, and D. Sun, "Cellular automata model of urban traffic emergency evacuation and rescue," *Journal of Traffic and Transportation Engineering*, vol. 11, no. 2, pp. 114–120, 2011.
- [7] J. C. Long, *Studies on Congestion Propagation Properties and Dissipation Control Strategies of Urban Road Traffic*, Beijing Jiaotong university, 2009.
- [8] Q. Chen and W. Wang, "Application of fuzzy optimization method based on CTM for traffic trunk line control under special events," *Journal of Southeast University (Natural Science Edition)*, vol. 38, no. 5, pp. 861–865, 2008.
- [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] A. P. Lian, Z. Y. Gao, and J. C. Long, "Dynamic user optimal assignment problem of link variables based on the cell transmission model," *Acta Automatica Sinica*, vol. 33, no. 8, pp. 852–859, 2007.

Research Article

Dynamic Recognition Model of Driver's Propensity under Multilane Traffic Environments

Xiaoyuan Wang,^{1,2} Jin Liu,¹ and Jinglei Zhang¹

¹ School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255091, China

² Department of Civil and Environmental Engineering, School of Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Correspondence should be addressed to Xiaoyuan Wang, wangxiaoyuan@sdut.edu.cn

Received 2 August 2012; Accepted 22 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Xiaoyuan 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.

Driver's propensity intends to change along with driving environment. In this paper, the situation factors (vehicle groups) that affect directly the driver's affection among environment factors are considered under two-lane conditions. Then dynamic recognition model of driver's propensity can be established in time-varying environment through Dynamic Bayesian Network (DBN). Physiology-psychology experiments and real vehicle tests are designed to collect characteristic data of driver's propensity in different situations. Results show that the model is adaptable to realize the dynamic recognition of driver's propensity type in multilane conditions, and it provides a theoretical basis for the realization of human-centered and personalized automobile active safety systems.

1. Introduction

With the rapid development of China economy, vehicle quantity, especially private vehicle, is increasing rapidly, and the contradiction among people, vehicle, and environment is increasing outstandingly in road traffic system. Above 90% of traffic accidents are caused by person, and above 70% of traffic accidents are caused by drivers. The reduction of traffic accidents not only needs to solve the problems of vehicle safety, road safety, and environment and climate impacts, but what is also more important is to research the influence of drivers on safe driving. Driver's propensity is a dynamic measurement of controller's affection, predilection, and others during driving. It is a core parameter to compute driver's intention and consciousness in safety driving assist systems, especially vehicle collision warning systems. Vehicle, as a mean of modern transportation, is convenient to people's

traveling; at the same time, it also brings some traffic safety problems. Automatic driving and driving assistant are vigorous and effective measures to reduce accidents and improve traffic safety. Driver's psychological and affective states are represented as driver's tendency [1] that is an important part of the driver-assistance systems, especially for the active security warning systems. Previous research about the driver's tendency focused mostly on the influence on traffic safety and the driver's psychological characteristics from relative static and macroscopic perspective [1–6]. Wang et al. [7–10] had researched preliminarily driver's tendency on special traffic scenes, such as free flow and car following; Feng and Fang et al. had researched cluster analysis of drivers' characteristics evaluation [11]; Chen et al. had researched subjective judgment of driving tenseness and control of vehicle motion [12]; Wang et al. had researched reliability and safety analysis methodology for identification of drivers' erroneous actions [13]; Cai and Lin had researched modeling of operators' emotion and task performance in a virtual driving environment [14]. However, they could not consider completely the influences of environment. In this paper, physiology-psychology experiments and real vehicle tests are designed to collect characteristic data of driver's propensity considering situation (vehicle group) that affects directly driver's affection among environment factors in different situations. Then dynamic recognition model of driver's propensity can be established in time-varying environment through Dynamic Bayesian Network. Results show that the model and relative experiment scheme are feasible. They can realize the dynamic recognition in multilane conditions.

2. Analysis of Traffic Situation Complexity

Vehicle group is crucial which consists of dynamic transport entity and its influence on driver's behaviors. Obviously, different vehicle position has different influence on target vehicle's driver. Within areas of influence, the front vehicle on the same lane has the largest effect on driver, then the around vehicles on the adjacent lanes, and rear vehicle on the same lane. The model can be simplified taking roads with two lanes in the same direction as an example and ignoring the influence from rear vehicle. The division of vehicle groups is shown in Figures 1 and 2.

Through simplifying the model further, the position of vehicles in left front, left side, left rear, right front, right side, and right rear can be represented into two types, limiting vehicles of left and right. When there is more than one limiting vehicle on target vehicle's left or right and the distance between them meets the minimum gap acceptance conditions, the vehicle whose spatial distance (distance between target and that vehicle along the direction of speed) is minimum can restrict target vehicle. If the distance does not meet the minimum gap acceptance conditions, then the two vehicles will be combined into one interference vehicle. So the complicated group is simplified as in Figure 3. Characteristics of driver's propensity for eight vehicle groups are shown in Table 1.

3. Dynamic Recognition Model of Driver's Propensity

3.1. Dynamic Bayesian Network

Dynamic Bayesian Network is also named Temporal Bayesian Network. It is static Bayesian Network developing with time. Every time slice corresponds to a Static Bayesian Network

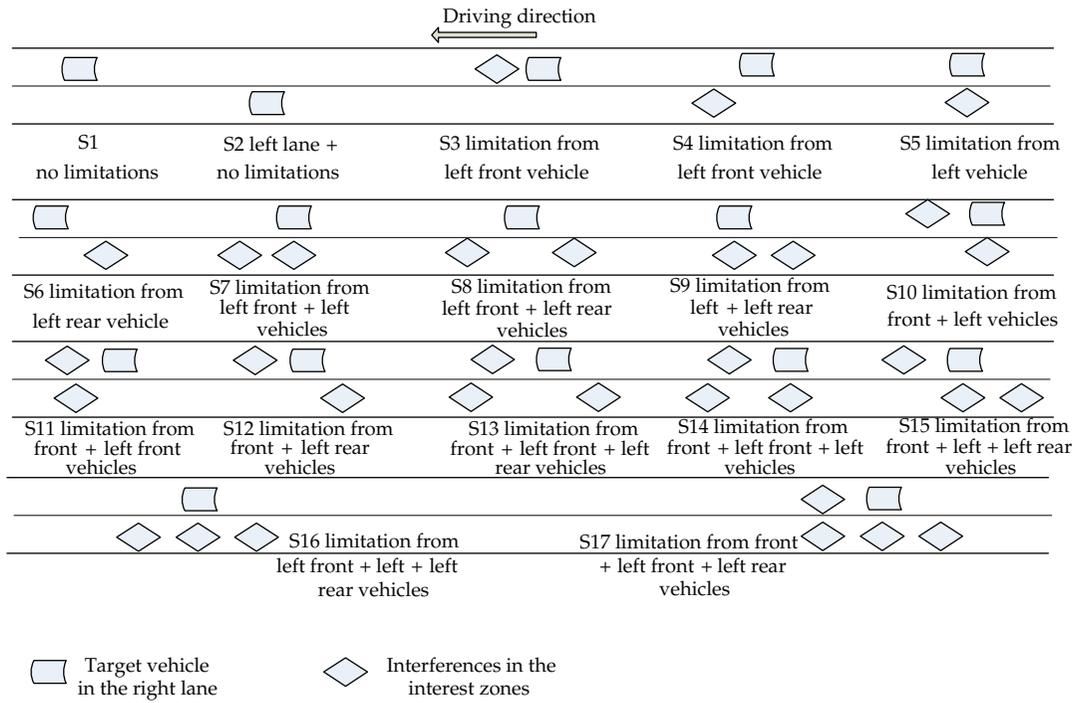


Figure 1: Vehicle group (A).

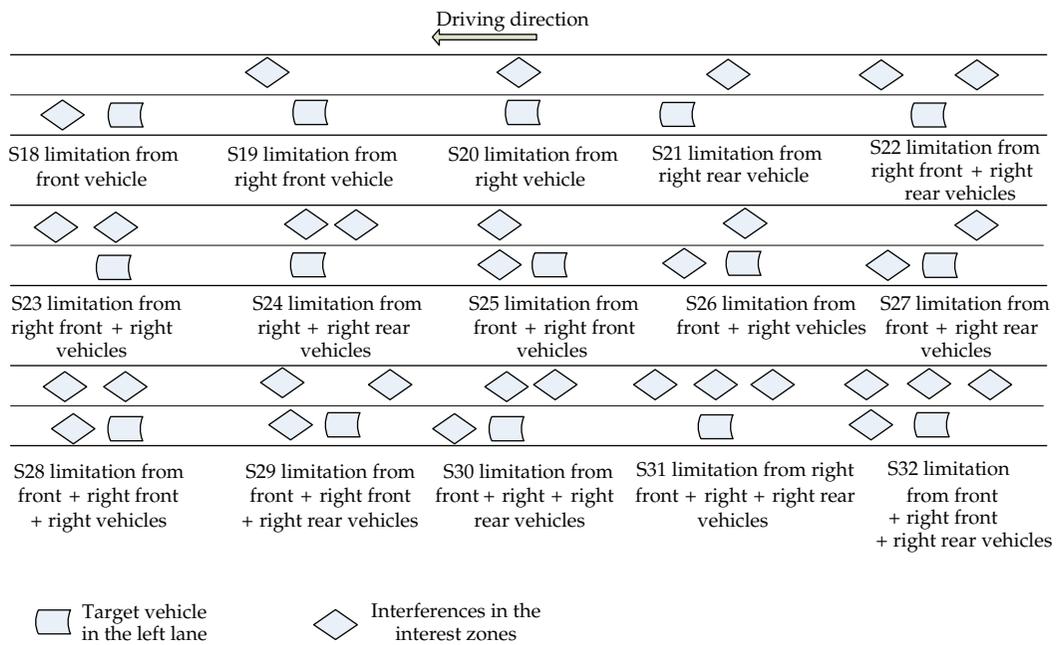
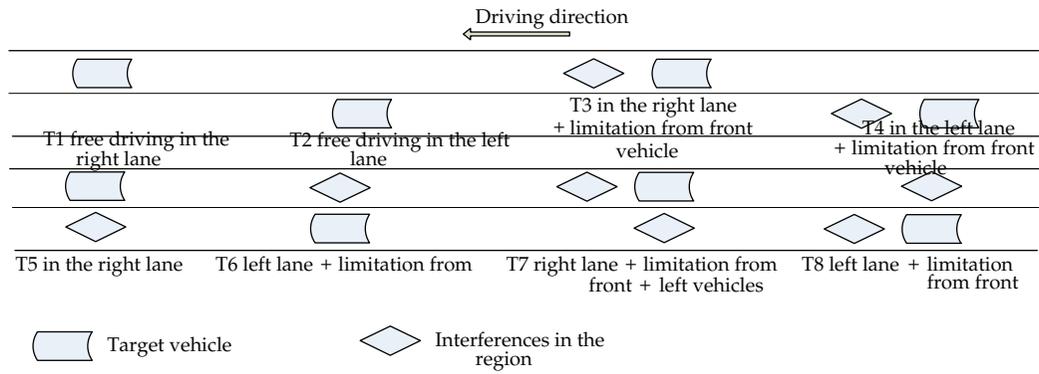


Figure 2: Vehicle group (B).

Table 1: Characteristics of driver's propensity for different groups.

Group	Characteristics of driver's propensity
T1 and T2	Speed of target vehicle; acceleration frequency; deceleration frequency; performance reaction time
T3 and T4	Speed of target vehicle; acceleration of target vehicle; headway; relative speed; relative acceleration; acceleration frequency; deceleration frequency; performance reaction time
T5 and T6	Speed of target vehicle; acceleration of target vehicle; acceleration frequency; deceleration frequency; risky lane-changing frequency; conservative lane-changing frequency; performance reaction time
T7 and T8	Headway; relative speed; acceleration frequency; deceleration frequency; risky lane-changing frequency; conservative lane-changing frequency; performance reaction time

**Figure 3:** Simplified vehicle group under two-lane conditions.

with identical structure and parameters. Two adjacent time slices are jointed by arc, which represents dependencies between adjacent time slices [15, 16].

Figure 4 shows a simple Dynamic Bayesian Network with three time slices, where, A_1 , A_2 , and A_3 are hide nodes; B_1 , B_2 , and B_3 are observed nodes. Each node is a variable. Variables have many states. Inference basis of Dynamic Bayesian Network is Bayes formula:

$$P(x | y) = \frac{P(yx)}{P(y)} = \frac{P(yx)}{\sum_x P(yx)}, \quad (3.1)$$

With n hide nodes and m observed nodes, inference essence of Static Bayesian Network is to calculate the following formula:

$$P(x_1, x_2, \dots, x_n | y_1, y_2, \dots, y_m) = \frac{\prod_j P(y_j | \text{Pare}(Y_j)) \prod_i P(x_i | \text{Pare}(X_i))}{\sum_{x_1, x_2, \dots, x_n} \prod_j P(y_j | \text{Pare}(Y_j)) \prod_i P(x_i | \text{Pare}(X_i))}, \quad (3.2)$$

$$i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m,$$

where x_i is a valued state of X_i , $\text{Pare}(Y_j)$ is parent node sets of Y_j , x_1, x_2, \dots, x_n located in below the \sum in the denominator is combination state of hide nodes, and \sum is the sum for

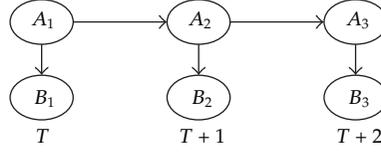


Figure 4: Simple Dynamic Bayesian Network.

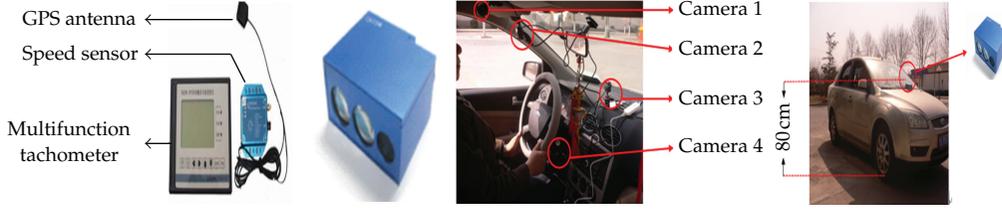


Figure 5: Dynamic human-vehicle-environment information acquisition systems.

joint distribution of observed variables and hidden variables combination state. In fact, it is to compute definite distribution of observed variables combination state.

Dynamic Bayesian Network consisted by T time slices can be obtained from Static Bayesian Network with time. Each time slice has n hide nodes and m observed nodes. Then inference of network can be expressed as follows:

$$\begin{aligned}
 & P(x_{11}, x_{12}, \dots, x_{1n}, \dots, x_{Tn} \mid y_{110}, y_{120}, \dots, y_{T10}, \dots, y_{Tm0}) \\
 &= \sum_{y_{11}y_{12}\dots y_{1m}\dots y_{T1}\dots y_{Tm}} \frac{\prod_{i,j} P(y_{ij} \mid \text{Pare}(Y_{ij})) \prod_{i,k} P(x_{ik} \mid \text{Pare}(X_{ik}))}{\sum_{x_{11}x_{12}\dots x_{1n}\dots x_{T1}\dots x_{Tn}} \prod_{i,j} P(y_{ij} \mid \text{Pare}(Y_{ij})) \prod_{i,k} P(x_{ik} \mid \text{Pare}(X_{ik}))} \quad (3.3) \\
 & \times \prod_{i,j} P(Y_{ij0} = y_{ij}), \quad i = 1, 2, \dots, T; j = 1, 2, \dots, m; k = 1, 2, \dots, n,
 \end{aligned}$$

where x_{ij} is a valued state of X_{ij} , i is the time slice of i , j is the hide node of j during the time slice of i , y_{ij} is the value of observed variable of Y_{ij} , $\text{Pare}(Y_{ij})$ is parent node sets of y_{ij} , Y_{ij0} is observed state of observed node j during time slice of i , and $P(Y_{ij0} = y_{ij})$ is the membership degree that continuous measurements of Y_{ij} belong to state y_{ij} .

3.2. Experiment Design

3.2.1. Experiment Equipment

The experiments designed in urban road environment collect data using dynamic human-vehicle-environment information acquisition systems (shown in Figure 5, including noncontact multifunction speedometer of SG299-GPS; laser range finder sensor of BTM300-905-200; high definition cameras; Minivap monitoring systems; HDTV camera; notebook computer.). Then driver's tendency can be extracted using the above data. In addition, the softwares used in the experiments include SPSS17.0 and Ulead VideoStudio10.0.

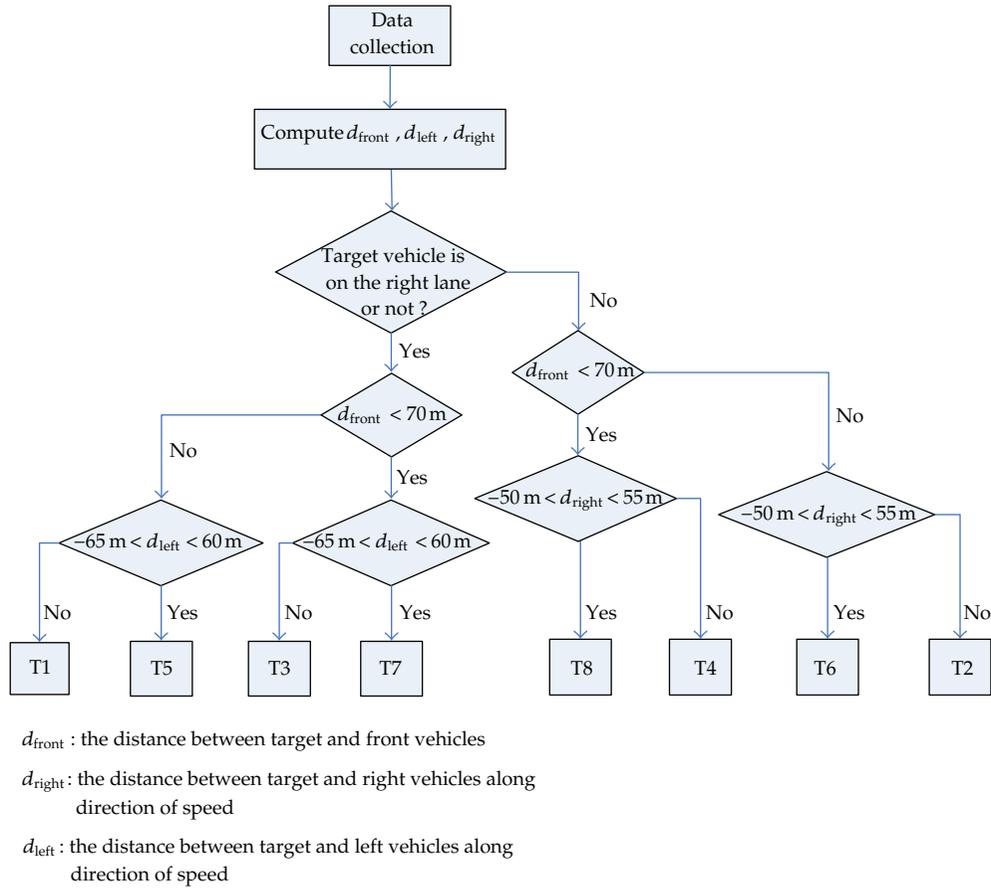


Figure 6: Recognition and identification model.

3.2.2. Experiment Conditions and Subjects

The experiments arranged in shiny days are taken from 8:00 am to 10:30 am on dry pavement, working day. Traffic is heavy, but there is no congestion. Sample capacity of experiment objects is 50, including 41 males and 9 females. Their ages range from 27 to 58 years old, average with 34.6 years. Driving years range from 3 to 22 years, with average 8.16 years.

3.2.3. Experiment Data

When human-vehicle-environment dynamic information is obtained, state division for the data is necessary to compute the membership degree in different states and to dynamically recognize driver's tendency. Computing model of state division and membership degree shown in chapter 2.3. Part of transited data is shown in Table 2.

Table 2: Typical data of drivers.

Type	Time	d_6	d_7	d_1	d_2	d_5	d_4	d_3
Ty1	1	(0.8, 0.1, 0.1)	(0.8, 0.1, 0.1)	(0.5, 0.3, 0.2)	(0.7, 0.2, 0.1)	(0.6, 0.2, 0.2)	(0.9, 0.1, 0.0)	(0.7, 0.2, 0.1)
	2	(0.8, 0.1, 0.1)	(0.7, 0.2, 0.1)	(0.6, 0.2, 0.2)	(0.7, 0.2, 0.1)	(0.6, 0.3, 0.1)	(0.9, 0.1, 0.0)	(0.7, 0.2, 0.1)
	3	(0.8, 0.1, 0.1)	(0.8, 0.1, 0.1)	(0.6, 0.3, 0.1)	(0.7, 0.2, 0.1)	(0.6, 0.3, 0.1)	(0.9, 0.1, 0.0)	(0.7, 0.2, 0.1)
Ty2	10
	1	(0.8, 0.1, 0.1)	(0.7, 0.2, 0.1)	(0.6, 0.2, 0.2)	(0.8, 0.1, 0.1)	(0.7, 0.2, 0.1)	(0.9, 0.1, 0.0)	(0.7, 0.2, 0.1)
	2	(0.4, 0.4, 0.2)	(0.6, 0.3, 0.1)	(0.4, 0.5, 0.1)	(0.5, 0.3, 0.2)	(0.5, 0.5, 0.0)	(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)
Ty3	1	(0.4, 0.4, 0.2)	(0.6, 0.3, 0.1)	(0.4, 0.5, 0.1)	(0.5, 0.3, 0.2)	(0.5, 0.5, 0.0)	(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)
	2	(0.4, 0.4, 0.2)	(0.6, 0.3, 0.1)	(0.4, 0.5, 0.1)	(0.5, 0.3, 0.2)	(0.5, 0.5, 0.0)	(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)
	3	(0.4, 0.4, 0.2)	(0.6, 0.3, 0.1)	(0.4, 0.5, 0.1)	(0.5, 0.3, 0.2)	(0.5, 0.4, 0.1)	(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)
Ty4	10
	1	(0.5, 0.4, 0.1)	(0.4, 0.4, 0.2)	(0.4, 0.5, 0.1)	(0.5, 0.3, 0.2)	(0.4, 0.4, 0.2)	(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)
	2	(0.2, 0.7, 0.1)	(0.2, 0.6, 0.2)	(0.2, 0.7, 0.1)	(0.2, 0.6, 0.2)	(0.2, 0.7, 0.1)	(0.1, 0.8, 0.1)	(0.2, 0.6, 0.2)
Ty5	1	(0.2, 0.6, 0.2)	(0.1, 0.7, 0.2)	(0.2, 0.7, 0.1)	(0.2, 0.7, 0.1)	(0.2, 0.7, 0.1)	(0.1, 0.8, 0.1)	(0.2, 0.6, 0.2)
	2	(0.2, 0.6, 0.2)	(0.2, 0.6, 0.2)	(0.2, 0.7, 0.1)	(0.2, 0.7, 0.1)	(0.2, 0.7, 0.1)	(0.1, 0.8, 0.1)	(0.2, 0.6, 0.2)
	3	(0.2, 0.6, 0.2)	(0.2, 0.6, 0.2)	(0.2, 0.7, 0.1)	(0.2, 0.7, 0.1)	(0.2, 0.7, 0.1)	(0.1, 0.8, 0.1)	(0.2, 0.6, 0.2)
Ty6	10
	1	(0.2, 0.7, 0.1)	(0.1, 0.8, 0.1)	(0.1, 0.6, 0.3)	(0.2, 0.6, 0.2)	(0.1, 0.7, 0.2)	(0.2, 0.7, 0.1)	(0.2, 0.6, 0.2)
	2	(0.1, 0.5, 0.4)	(0.2, 0.4, 0.4)	(0.1, 0.4, 0.5)	(0.1, 0.3, 0.6)	(0.2, 0.4, 0.4)	(0.1, 0.4, 0.5)	(0.2, 0.4, 0.4)
Ty7	1	(0.1, 0.5, 0.4)	(0.2, 0.4, 0.4)	(0.1, 0.4, 0.5)	(0.2, 0.4, 0.4)	(0.2, 0.4, 0.4)	(0.1, 0.4, 0.5)	(0.2, 0.4, 0.4)
	2	(0.1, 0.5, 0.4)	(0.2, 0.4, 0.4)	(0.1, 0.4, 0.5)	(0.2, 0.4, 0.4)	(0.2, 0.4, 0.4)	(0.1, 0.4, 0.5)	(0.2, 0.4, 0.4)
	3	(0.1, 0.5, 0.4)	(0.2, 0.4, 0.4)	(0.1, 0.4, 0.5)	(0.2, 0.4, 0.4)	(0.1, 0.4, 0.5)	(0.1, 0.4, 0.5)	(0.2, 0.4, 0.4)
Ty8	10
	1	(0.1, 0.4, 0.5)	(0.1, 0.4, 0.5)	(0.1, 0.4, 0.5)	(0.1, 0.4, 0.5)	(0.1, 0.4, 0.5)	(0.1, 0.4, 0.5)	(0.2, 0.4, 0.4)
	2	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)
Ty9	1	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)
	2	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)
	3	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)
Ty10	10
	1	(0.1, 0.2, 0.7)	(0.0, 0.2, 0.8)	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)
	2	(0.1, 0.2, 0.7)	(0.0, 0.2, 0.8)	(0.1, 0.2, 0.7)	(0.1, 0.1, 0.8)	(0.1, 0.1, 0.8)	(0.2, 0.2, 0.6)	(0.1, 0.2, 0.7)

Note: Ty1 is conservative type; Ty2 is common-conservative type; Ty3 is common type; Ty4 is common-radical type; Ty5 is radical type. d_1 is acceleration frequency, d_2 is deceleration frequency, d_3 is performance reaction time, d_4 is risky lane-changing frequency, d_5 is conservative lane-changing frequency, d_6 is headway, and d_7 is relative speed.

Table 3: Conditional probability matrices of driver's characteristics under T7 conditions.

Type	Conditional probability matrices of driver's characteristics															
	$P(d_1 \text{human})$ (small, medium, and large)		$P(d_2 \text{human})$ (large, medium, and small)		$P(d_3 \text{human})$ (slow, moderate, and fast)		$P(d_4 \text{human})$ (high, medium, and low)		$P(d_5 \text{human})$ (low, medium, and high)							
Ty1	0.75	0.15	0.10	0.85	0.10	0.05	0.80	0.10	0.10	0.05	0.90	0.05	0.05	0.80	0.10	0.10
Ty2	0.45	0.45	0.10	0.40	0.45	0.15	0.50	0.40	0.10	0.30	0.60	0.30	0.10	0.10	0.30	0.10
Ty3	0.15	0.75	0.10	0.10	0.80	0.10	0.10	0.80	0.10	0.70	0.15	0.70	0.15	0.10	0.80	0.10
Ty4	0.10	0.40	0.50	0.10	0.45	0.45	0.10	0.45	0.45	0.25	0.15	0.25	0.60	0.10	0.40	0.50
Ty5	0.10	0.10	0.80	0.10	0.15	0.75	0.10	0.10	0.80	0.15	0.10	0.15	0.75	0.05	0.15	0.80

Notes: The above table shows the corresponding probability when driver's characteristics are conservative type, common-conservative type, common type, common-radical type and radical type. For example, when driver's characteristics are conservative type, the small probability of d_{12} is 75%, the medium probability is 15%, and the large probability is 10%.

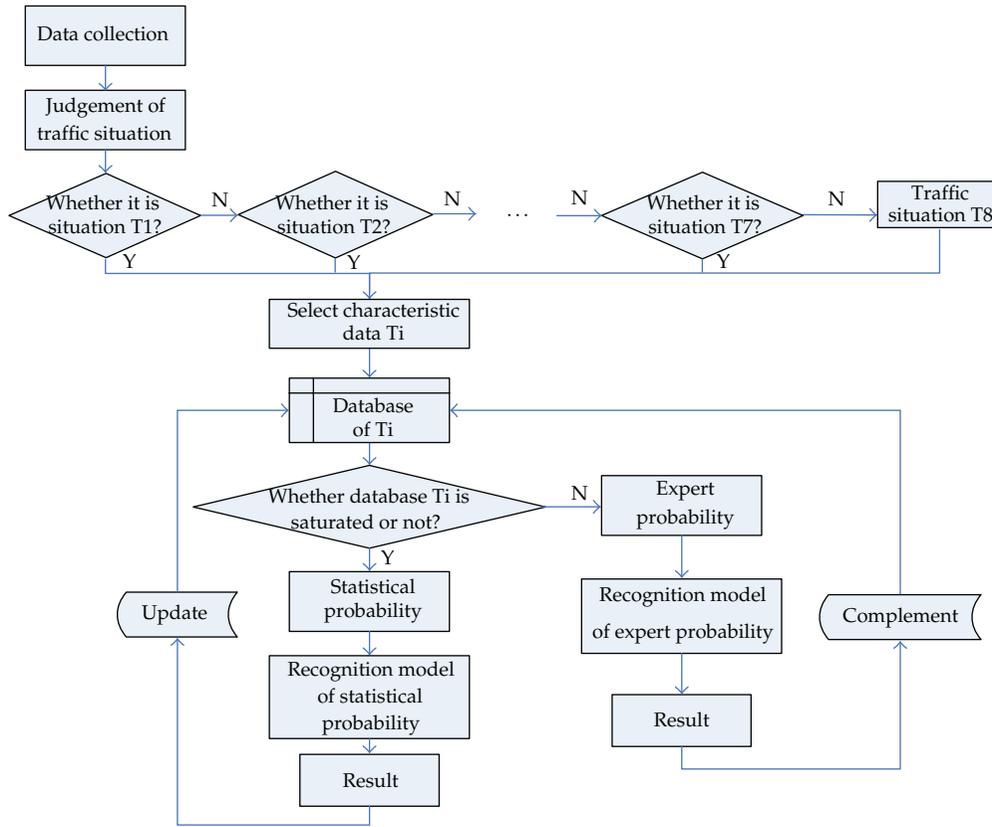


Figure 7: Flow chart of the model.

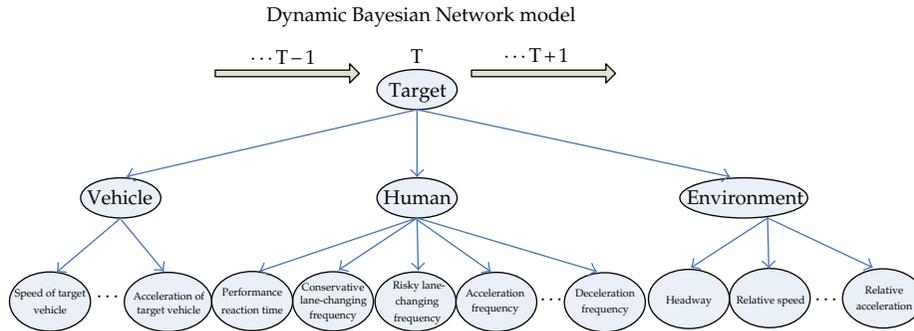


Figure 8: Model of Dynamic Bayesian Network (union of characteristic data).

3.3. Recognition of Driver's Propensity Based on Dynamic Bayesian Network

Vehicle group recognition depends on a group of different space location and distance along with direction of speed. Amounts of experiments show that when d_{left} or d_{right} (d_{left} is space line distance between target vehicle and vehicles in the left lane along with direction of target vehicle's speed, and d_{right} is space line distance between target vehicle and vehicles in the right lane along with direction of target vehicle's speed) are less than (or equal to) one

Table 4: Conditional probability matrices of environmental characteristics under T7 conditions.

Type	Conditional probability matrices of environmental characteristics					
	$P(d_6 \text{environment})$ (large, medium, and small)			$P(d_7 \text{environment})$ (small, medium, and large)		
Ty1	0.80	0.10	0.10	0.75	0.15	0.10
Ty2	0.50	0.40	0.10	0.50	0.45	0.05
Ty3	0.10	0.80	0.10	0.10	0.80	0.10
Ty4	0.10	0.45	0.45	0.10	0.40	0.50
Ty5	0.05	0.10	0.85	0.05	0.15	0.80

Table 5: Conditional probability matrices of tendency type under T7 conditions.

Type	Conditional probability matrices of driver's propensity type									
	$P(\text{human factor} \text{propensity})$					$P(\text{environment} \text{propensity})$				
	(Ty1, Ty2, Ty3, Ty4, and Ty5)					(Ty1, Ty2, Ty3, Ty4, and Ty5)				
Ty1	0.75	0.10	0.05	0.05	0.05	0.65	0.15	0.10	0.05	0.05
Ty2	0.15	0.65	0.10	0.05	0.05	0.10	0.70	0.10	0.05	0.05
Ty3	0.05	0.10	0.70	0.10	0.05	0.05	0.10	0.70	0.10	0.05
Ty4	0.05	0.05	0.10	0.70	0.10	0.05	0.05	0.10	0.65	0.15
Ty5	0.05	0.05	0.10	0.15	0.65	0.05	0.10	0.10	0.15	0.60

Notes: the above table shows the corresponding probability when driver's or environment characteristics are conservative type, common-conservative type, common type, common-radical type, and radical type. For example, when driver's propensity is conservative type, the probability that driver's characteristics belong to conservative type is 75%, belong to common and conservative type is 10%, belong to common type is 5%, belongs to common and radical type is 5%, and belong to radical type is 5%.

threshold value, it will affect the target vehicle. Amounts of data for different drivers show that the interval of d_{left} is $(-65 \text{ m}, 60 \text{ m})$ and interval of d_{right} is $(-50 \text{ m}, 55 \text{ m})$. Recognition and identification model of vehicle group is shown in Figure 6. Flow chart of the model for Dynamic Bayesian Network is shown in Figure 7. Model of Dynamic Bayesian Network is shown in Figure 8.

Figure 8 contains all characteristic data in different groups. According to different environments and corresponding characteristic data, computing can be made in the process of recognition and identification. Variable state sets in Dynamic Bayesian Network are shown as follows.

Driver's propensity includes conservative type, common-conservative type, common type, common-radical type, and radical type; speed of target vehicle includes small, medium, and large; acceleration of target vehicle includes small, medium, and large; headway includes large, medium, and small; relative speed includes slow, moderate, and fast; relative acceleration includes small, medium, and large; deceleration frequency includes high, middle, and low; acceleration frequency includes high, middle, and low; performance reaction time includes long, medium, and short; conservative lane-changing frequency includes high, middle, and low; risky lane-changing frequency includes low, middle, and high.

Variable states are fuzzy set. Definition of state is derived from relative change of data during driving. If states are divided uniformly, then the differences of driver's characters cannot be represented truly. State thresholds of drivers are different. The data of inputting model is expressed with probability. In this paper, membership degree is to express the probability of certain characteristic data. If sample data x contains N characteristic data, it

Table 6: State transition probability of Dynamic Bayesian Network under T7 conditions.

New/old node	State transition probability of Dynamic Bayesian Network				
	Ty1 (old)	Ty2 (old)	Ty3 (old)	Ty4 (old)	Ty5 (old)
Ty1 (new)	0.60	0.20	0.10	0.05	0.05
Ty2 (new)	0.15	0.60	0.15	0.05	0.05
Ty3 (new)	0.05	0.10	0.70	0.10	0.05
Ty4 (new)	0.05	0.05	0.15	0.60	0.15
Ty5 (new)	0.05	0.05	0.10	0.15	0.65

Table 7: Initial probability of different driver's propensity.

Calibration result	Initial probability				
	P (Ty1)	P (Ty2)	P (Ty3)	P (Ty4)	P (Ty5)
Ty1	0.75	0.10	0.05	0.05	0.05
Ty2	0.10	0.70	0.10	0.05	0.05
Ty3	0.05	0.10	0.70	0.10	0.05
Ty4	0.05	0.05	0.10	0.70	0.10
Ty5	0.05	0.05	0.05	0.10	0.70

will be expressed with the value of membership degree. P_i is probability that characteristic component is subordinate to i . There are three kinds of eigenvector state. Calculation formula of membership degree is shown as follows:

$$\begin{aligned}
 P_1 &= \left(1 + \left| \frac{a_i - a_{i \min}}{a_{i \max} - a_i} \right| \right)^{-4}, \\
 P_2 &= \left(1 + \left| \frac{a_i - \bar{a}_i}{a_{i \max} - a_i} \right| \right)^{-1}, \\
 P_3 &= 1 - P_1 - P_2,
 \end{aligned} \tag{3.4}$$

where \bar{a}_i is mean value of known sample data, a_i is observed value of characteristic data, and $a_{i \min}$ and $a_{i \max}$ are minimum and maximum of observed values.

3.3.1. Prophase Parameter Setting

Conditional probability matrix is a kind of expert knowledge, which represents an opinion of causality between correlative nodes in network. According to expert experience, characteristic data of driver's propensity includes headway, relative speed, deceleration frequency, acceleration frequency, performance reaction time, conservative lane-changing frequency, and risky lane-changing frequency during stable driving under vehicle group of T7. So its inference rule is probabilistic manner. Initial conditional probability is got by expert experiences. When the number of data in database reaches to a certain capacity, probability will be got by computing.

According to the above inference rule, conditional probability matrices of driver's characteristics are gained and shown in Tables 3, 4, 5, and 6. d_1 is acceleration frequency, d_2 is

Table 8: Recognition and identification result of driver's propensity (expert probability).

Type	Time	Recognition and identification result (expert probability)				
		Ty1	Ty2	Ty3	Ty4	Ty5
Ty1	1	0.63067	0.14054	0.10462	0.06543	0.05874
	2	0.66285	0.15179	0.09091	0.04815	0.0463
	3	0.68018	0.14084	0.11021	0.01992	0.04984

	10	0.85191	0.06092	0.03991	0.03909	0.00817
Ty2	1	0.14124	0.61423	0.12424	0.07676	0.04353
	2	0.12077	0.64089	0.13091	0.05914	0.04829
	3	0.13541	0.65434	0.11333	0.05559	0.04133

	10	0.05357	0.84156	0.06255	0.01744	0.02488
Ty3	1	0.04137	0.15113	0.61122	0.12872	0.06756
	2	0.05448	0.14411	0.64415	0.13575	0.02151
	3	0.05333	0.12483	0.68329	0.10621	0.03234

	10	0.04184	0.05131	0.84024	0.05879	0.00782
Ty4	1	0.01266	0.06271	0.11206	0.64751	0.16506
	2	0.03132	0.03185	0.12166	0.65865	0.15652
	3	0.03257	0.06239	0.10242	0.66757	0.13505

	10	0.02226	0.02633	0.04422	0.85572	0.05147
Ty5	1	0.06148	0.06211	0.09043	0.17782	0.60816
	2	0.04468	0.07452	0.08471	0.15531	0.64078
	3	0.04462	0.03474	0.09399	0.14577	0.68088

	10	0.02312	0.02257	0.05262	0.04704	0.85465

deceleration frequency, d_3 is performance reaction time, d_4 is risky lane-changing frequency, d_5 is conservative lane-changing frequency, d_6 is headway, and d_7 is relative speed.

It is noticed that conditional probability matrix is a kind of expert knowledge, so it has certain subjectivity. Sample data can be debugged repeatedly. Matrix data can be adjusted reasonably to improve the creditability of assessment result.

Due to the limited space in this paper, prophase parameter setting of other groups is not amplified any more.

3.3.2. Anaphase Parameter Setting

When the number of data reaches to a certain capacity, the database of driver's propensity can be established. According to driver's psychology test results, the data is classified to five types: conservative type, common-conservative type, common type, common-radical type, and radical type. Data of each type consist of characteristic data extracted and the recognition result of driver's propensity in prophase stage. In the same results of psychology tests, statistical analysis for data is to determine the proportion of characteristic data from

Table 9: Recognition and identification result of driver's propensity (statistical probability).

Type	Time	Recognition and identification result (statistical probability)				
		Ty1	Ty2	Ty3	Ty4	Ty5
Ty1	1	0.73461	0.15486	0.06825	0.03038	0.01190
	2	0.77525	0.11735	0.05814	0.03394	0.01532
	3	0.81994	0.11206	0.01064	0.04224	0.01513

	10	0.93704	0.02811	0.01651	0.01004	0.00829
Ty2	1	0.10751	0.71301	0.11708	0.02018	0.04223
	2	0.09211	0.72634	0.09094	0.04531	0.04530
	3	0.07130	0.81076	0.08220	0.02817	0.00758

	10	0.02198	0.94139	0.02289	0.01247	0.00128
Ty3	1	0.06365	0.11161	0.70473	0.11451	0.00550
	2	0.02781	0.11532	0.70561	0.11371	0.03755
	3	0.03796	0.10864	0.74603	0.10353	0.00384

	10	0.00302	0.03251	0.94079	0.02103	0.00264
Ty4	1	0.01647	0.05396	0.09593	0.73192	0.10172
	2	0.00506	0.02362	0.10056	0.76312	0.10764
	3	0.01482	0.03376	0.08892	0.79106	0.07145

	10	0.00227	0.01161	0.02368	0.93504	0.02741
Ty5	1	0.02530	0.03231	0.07193	0.16028	0.71018
	2	0.02938	0.02224	0.07548	0.14080	0.73210
	3	0.01952	0.02803	0.06931	0.11413	0.76901

	10	0.00022	0.00671	0.01712	0.03149	0.94446

driver's propensity in different traffic environments in order to determine the conditional probabilities in Dynamic Bayesian Network. The determination of state transition probability of Dynamic Bayesian Network is similar to that of conditional probability, so the process is not amplified any longer.

3.4. Model Verification

There are two parts of recognition and identification model. Firstly, recognition can be taken with data from expert experiences. Secondly, recognition of driver's tendency can be taken with statistical data. In the situation of absence of another evidence, initial states depend on initial value set with driver's propensity, which is shown in Table 7.

According to the above several circumstances, initial values of different drivers are taken as a rational starting point. Evidence in different nodes can be collected (assuming independent). Vast characteristic data and recognition results for several drivers can be collected in this paper. The data of five typical driver's propensity (initial calibration) is

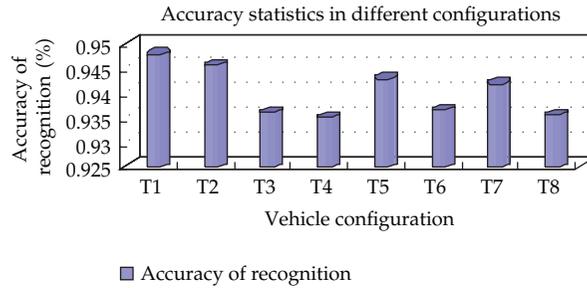


Figure 9: Accuracy of recognition in different situations.

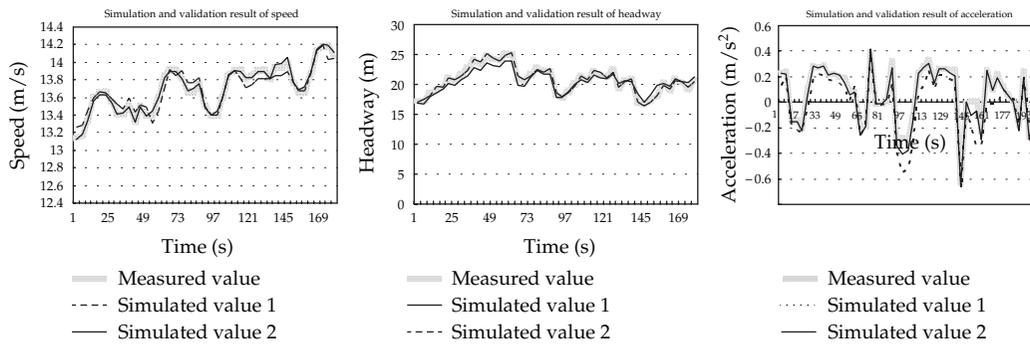


Figure 10: Verification results of speed, headway, and acceleration.

amplified under T7 conditions. Tables 8 and 9 are recognition and identification results of driver’s propensity (includes expert probability and statistical probability).

The same method is used to verify accuracy of recognition for driver’s propensity in different groups. The result is shown in Figure 9.

Verification results are shown in Figure 10. Curve 1 is the result without considering the change of driver’s propensity in the simulation process. Curve 2 shows the situation process with considering driver’s propensity in real time.

Microscopic models considering differences of driver’s propensity are more precise to simulate driver’s behaviors. Meanwhile, scope of application is very broad.

Accuracy of recognition and identification model is relatively higher under multilane environments. It also can meet the need of dynamic recognition for driver’s propensity under multilane conditions.

4. Conclusion

Driver’s propensity can represent their affection states in the process of vehicle operation and movement. It can change along with environment and affect profoundly drivers’ cognition and process procedure on environment information. Therefore, the real-time identification of driver’s state is the key to realize the driver-assistance systems and the active security warning systems. In this paper, situation factors (vehicle group) that affect directly driver’s affection among environment factors are considered under two-lane conditions. Then dynamic recognition and identification model of driver’s propensity can be established

in time-varying environment through Dynamic Bayesian Network. It also can provide a theoretical basis for the realization of human-centered and personalized automobile active safety systems. For three-lane or more complicated environments, recognition and computing of driver's affection need further research.

Acknowledgments

This project is supported by National Natural Science Foundation of China (61074140), Natural Science Foundation of Shandong Province (ZR2010FM007 and ZR2011EEM034), Key Disciplines (Lab) Excellent Skeleton Teachers International Exchange Visitor Program of Shandong Province, and Young Teacher Development Support Project of Shandong University of Technology.

References

- [1] X. Q. Feng and C. Q. Fang, "Cluster analysis of evaluation index on driver's characteristics," *Communications science and technology Heilongjiang*, no. 11, pp. 161–163, 2007.
- [2] B. G. Hillel and D. Shinar, "The tendency of drivers to pass other vehicles," *Transportation Research F*, vol. 8, no. 6, pp. 429–439, 2005.
- [3] C. Pêcher, C. Lemerrier, and J. M. Cellier, "Emotions drive attention: effects on driver's behaviour," *Safety Science*, vol. 47, no. 9, pp. 1254–1259, 2009.
- [4] O. T. Ben-Ari and E. Shay, "The association between risky driver and pedestrian behaviors: the case of Ultra-Orthodox Jewish road users," *Transportation Research F*, vol. 15, no. 2, pp. 188–195, 2012.
- [5] I. A. Kaysi and A. S. Abbany, "Modeling aggressive driver behavior at unsignalized intersections," *Accident Analysis and Prevention*, vol. 39, no. 4, pp. 671–678, 2007.
- [6] H. W. Warner, T. Özkan, T. Lajunen, and G. Tzamalouka, "Cross-cultural comparison of drivers' tendency to commit different aberrant driving behaviours," *Transportation Research F*, vol. 14, no. 5, pp. 390–399, 2011.
- [7] X. Y. Wang and J. L. Zhang, "Extraction and recognition methods of vehicle driving tendency feature based on driver-vehicle-environment dynamic data under car following," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1269–1281, 2011.
- [8] J. L. Zhang and X. Y. Wang, "The feature extraction and dynamic deduction method of vehicle driving tendency under time variable free flow condition," *Journal of Beijing Institute of Technology*, vol. 20, no. 1, pp. 127–133, 2011.
- [9] Y. Y. Zhang, X. Y. Wang, and J. L. Zhang, "Verification method of vehicle driving tendency recognition model under free flow," *Journal of Computer Application*, vol. 32, no. 2, pp. 578–580, 2012.
- [10] Y. Y. Zhang, X. Y. Wang, and J. L. Zhang, "The verification method of vehicle driving tendency recognition model based on driving simulated experiments under free flow," *Journal of Wuhan University of Technology*, vol. 33, no. 9, pp. 82–86, 2011.
- [11] Y. Q. Feng, C. Q. Fang et al., "Cluster analysis of drivers characteristics evaluation," *Communications Science and Technology Heilongjiang*, no. 11, pp. 161–163, 2007.
- [12] X. M. Chen, L. Gao, and S. B. Wu, "Research on subjective judgment of driving tenseness and control of vehicle motion," *Journal of Highway and Transportation Research and Development*, vol. 24, no. 8, pp. 144–148, 2007.
- [13] 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.
- [14] H. Cai and Y. Lin, "Modeling of operators emotion and task performance in a virtual driving environment," *International Journal of Human Computer Studies*, vol. 69, no. 9, pp. 571–586, 2011.
- [15] Z. B. Zhou, C. Q. Ma, J. L. Zhou, and D. D. Dong, "Dynamic fault tree analysis based on dynamic Bayesian networks," *Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice*, vol. 28, no. 2, pp. 35–42, 2008.
- [16] S. Gao, Q. Pan, Z. J. Li, and Q. K. Xiao, "Optimal state estimation of moving targets based on DBN," *Journal of Detection and Control*, vol. 29, no. 4, pp. 74–76, 2007.

Research Article

Distinguishing between Rural and Urban Road Segment Traffic Safety Based on Zero-Inflated Negative Binomial Regression Models

Xuedong Yan,¹ Bin Wang,¹ Meiwu An,² and Cuiping Zhang¹

¹ MOE Key Laboratory for Transportation Complex Systems Theory and Technology,
School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

² Saint Louis County Highways and Traffic and Public Works, 121 South Meramec Avenue,
Saint Louis, MO 63105, China

Correspondence should be addressed to Xuedong Yan, xdyan@bjtu.edu.cn

Received 3 August 2012; Accepted 15 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Xuedong 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.

In this study, the traffic crash rate, total crash frequency, and injury and fatal crash frequency were taken into consideration for distinguishing between rural and urban road segment safety. The GIS-based crash data during four and half years in Pikes Peak Area, US were applied for the analyses. The comparative statistical results show that the crash rates in rural segments are consistently lower than urban segments. Further, the regression results based on Zero-Inflated Negative Binomial (ZINB) regression models indicate that the urban areas have a higher crash risk in terms of both total crash frequency and injury and fatal crash frequency, compared to rural areas. Additionally, it is found that crash frequencies increase as traffic volume and segment length increase, though the higher traffic volume lower the likelihood of severe crash occurrence; compared to 2-lane roads, the 4-lane roads have lower crash frequencies but have a higher probability of severe crash occurrence; and better road facilities with higher free flow speed can benefit from high standard design feature thus resulting in a lower total crash frequency, but they cannot mitigate the severe crash risk.

1. Introduction

Previous studies have been focused on distinguishing between rural and urban traffic safety using traffic crash data, but the influence of rural or urban settings on segment safety is controversial. The fatal traffic crash research indicated that fatality rates in rural areas are higher than in urban areas [1–3]. The higher fatality and injury rates in rural road facilities

have been attributed to various reasons, such as longer emergency response time and further distance to crash locations [3]. The other explanations include higher speed limit and worse driving habits (e.g., alcohol, a lower rate of seat belt use, and safety precaution use), road conditions, and accessibility to trauma care [4–6]. On the other hand, some studies reported that crash frequencies in urban areas were higher than those in rural areas [7, 8], because urban regions involve more complex traffic conditions, high traffic volume, congestion, poor pavement conditions, and so forth [9]. While the risk of severe crashes appears higher in rural segments, no differences were identified in the cause of injury and place of injury between urban and rural drivers [10].

Neither crash frequencies nor fatality rate can entirely represent the influence of rural or urban settings on the segments. Many other factors lead to the occurrence of traffic crashes, such as traffic characteristics, road design characteristics, demographic features, and pavement maintenance conditions [11–17]. Therefore, numerous cross-sectional studies have been conducted to characterize the relationships between factors and road segment-involved crashes. In the prior studies, the Poisson models are the most common ones which have been widely used [18, 19]. It is known that a Poisson model would be appropriate only when the mean and the variance of the crash frequencies are approximately equal. However, this assumption has been proved invalid for modeling traffic crash frequencies [20, 21], because the variances of crash frequencies were generally greater than means. Therefore, the negative binomial (NB) regression models were introduced to overcome this overdispersion problem, which had a more flexible mean-variance equality constraint [16, 22]. Nevertheless, both Poisson and NB models cannot deal with the property of crash frequency data with a large density of zeros (no crashes occur on roads during the observation period). Correspondingly, the zero-inflated count regress models were developed and applied for analyzing and predicting crash frequencies. The zero-inflated count regression models are capable of handling the apparent “excess” zeros crash data and generally have a more statistical suitability for modeling crash data than Poisson and NB regression models [23].

In order to understand the role of rural or urban settings in segment safety, the crash rate, crash frequencies, and the injury and fatality frequencies are taken into consideration in this study to distinguish between rural and urban traffic safety. The GIS-based crash data during four and half years in Pikes Peak Area, USA were applied for analyses. The GIS techniques for traffic data process have been proved effective to analyze and visualize crash data [24] and have advantages in data display, clear presentation of spatial relationship, and convenient query of relevant data [25, 26]. Since previous studies have discussed the suitability of various models in the prediction of crash frequencies, we adopt zero-inflated negative binomial (ZINB) regression models for crash frequency analysis and prediction, because zero-crash segments account for more than 40% of the total data in this study.

2. Methodology

2.1. Data Preparation

Accident data were obtained from the department of revenue (DOR) and were calculated by the total accidents recorded from 2006 July to 2010 December. It contains useful traffic information, such as crash location, severity, weather condition, and segment type, and the data were geocoded into GIS databases by the PPACG (Pikes Peak Area Council of Governments).

Based on the GIS process of spatial join between whole road network and urban boundary, the road segments were classified into two categories: rural segments and urban

Table 1: ZINB models for segment crash frequency analyses and predictions.

Role	Variables	Types	Description
Dependent variable	Num_crsh	Continuous	Number of crashes (crash frequencies)
	ADT_1000	Continuous	Average thousand daily traffic (thousand vehicles per day)
	Length	Continuous	The length of a segment (mile)
Independent variable	FFS	Continuous	Free flow speed in road segment (mph)
	Daily_VMT	Continuous	Average daily Vehicle Mile Traveled (VMT)
	RorU	Discrete	Rural or urban: 0 = urban; 1 = rural
	FC	Discrete	Segment functional classification: 0 = interstate; 1 = expressway; 2 = arterial; 3 = M_Arterial
	Numberofla	Discrete	Number of lanes: 0 = 2 lanes; 1 = 4 lanes

segments. Before analyzing segment crashes, the crashes at intersections were separated from the databases. Thus, the 200-ft intersection buffers were first created, and the crashes within these intersection buffers were deleted from the segment crash analyses. Then, with a road-segment layer separated from the road network geodatabase, the crashes associated with segments needed to be further separated from all other crashes. Because these segments may have wide cross-sections, a 150-foot buffer on both sides of an arterial centerline was adopted to capture most crashes associated with the segments only. After the 150 foot buffers were created, the crashes within these buffers were selected and aggregated in their corresponding segments.

Because different categories of road facilities vary by characteristics of highway design, traffic operation, and environments, the crash data associated with a specific type of highways needed to be separated from the other types of highways. In this study, the crash risk was calculated and analyzed not only for the overall segment network, but also for interstate, expressway, principal arterial, and minor arterial, respectively. The segments belonging to other road types were excluded from these segments. The combined data set was further organized according to the following criteria.

- (i) These accidents were divided into three categories: fatal, injury, and property-damage only (PDO) accounting for the accident severity.
- (ii) Road segments with 2 and 4 lanes were selected, because 6 lanes segments exist in urban areas only.
- (iii) ADT was calculated by 1000, because the change in crash frequency with increment of one vehicle is meaningless.

The cleaned accident data were overlaid with the GIS-based network and distributed into each segment in rural and urban areas. The segments were first analyzed and compared in terms of crash rate based on the comparative statistics of the four types of road segments. Then, ZINB models for segment crash frequency analyses and predictions were developed, in which variables are described in Table 1.

2.1.1. Zero-Inflated Negative Binomial Regression

For a Poisson crash frequency model, it assumes that the observed crash count data y_i , given the vector of covariate x_i , follows a Poisson distribution. The density function of y_i can be expressed as follows:

$$P(Y_i = y_i | x_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots, \quad (2.1)$$

where the parameter u_i , conditional mean number of events for each covariate x_i , is given by

$$\mu_i = \text{Exp}(x_i' \beta), \quad (2.2)$$

where β is a $(k+1) \times 1$ parameter vector (β_0 is the coefficient for intercept, and $\beta_1, \beta_2, \dots, \beta_k$ are for k regressors).

In the Poisson regression, the conditional variance of the count variable is equal to the conditional mean as follows:

$$V(y_i | x_i) = E(y_i | x_i) = \mu_i, \quad (2.3)$$

where x_i is the covariate of road segment geometric and traffic features in each record including the intercept; u_i is the conditional mean of the crash frequency y_i . Since this assumption is contradict to the fact that the vehicle accident data are always significantly overdispersed relative to its mean, the NB regression model was developed with a heterogeneity component accounting for unobserved heterogeneity in the crash count data as follows:

$$u_i = \text{Exp}(x_i \beta + \varepsilon_i) = \text{Exp}(x_i \beta) \text{Exp}(\varepsilon_i), \quad (2.4)$$

where β is the parameter coefficients vector to be estimated for independent variables including intercept; $\text{Exp}(\varepsilon_i)$ is a heterogeneity component accounting for unobserved heterogeneity in the crash count data, which is independent of x_i . However, there is always a large density of zeros in crash count data, which cannot accurately be predicted by traditional NB models. For this situation, the zero-inflated regression models were developed in the crash frequency-related research area.

Zero-inflated count models provide a way of modeling the excess zeros in addition to allowing for overdispersion. For each road segment, there are two possible data generation processes. Process 1 is chosen with probability ω_i and process 2 with probability $1-\omega_i$. Process 1 generates only zero counts, whereas process 2 generates counts from either a poisson or a negative binomial model. In this paper, the probability ω_i depends on the geometric and traffic features of segment i , can be obtained from the logistic function F , as follows:

$$\omega_i = F(z_i' \gamma) = \Lambda(z_i' \gamma) = \frac{\exp(z_i' \gamma)}{1 + \exp(z_i' \gamma)}, \quad (2.5)$$

Table 2: Original statistics for the length and mileage.

Area	Number	Mean length	Minimum	Maximum	Std. deviation	Mileage
Rural	374	0.968	0.063	5.998	0.989	362.100
Urban	2387	0.301	0.051	3.992	0.382	719.239
Network	2761	0.392	0.051	5.998	0.525	1081.339

where z'_i is the vector of independent variables specified in the logistic regression model (road facility and traffic features) and intercept; γ is the vector of zero-inflated coefficients to be estimated.

The probability of crash frequency for segment i can be expressed as follows:

$$p(y_i | x_i, z_i) = \begin{cases} \omega_i + (1 - \omega_i)g(y_i | x_i), & y_i = 0, \\ (1 - \omega_i)g(y_i | x_i), & y_i > 0, \end{cases} \quad (2.6)$$

where $g(y_i | x_i)$ follows either Poisson distribution or NB distribution; x_i is the vector of covariates of observation i specified in the model.

In this study, ZINB models were used for regression efforts because zero-crash segments account for more than 40% of the total data.

3. Results

3.1. Comparative Statistical Analyses of Rural and Urban Traffic Safety

During the observation period of four and a half years, there were 9651 crashes occurring in the study areas, consisting of 1057 records in rural segments and 8594 records in urban segments. Among the crashes in the rural segments, there were 15 fatal and 176 injured accidents. On the other hand, 46 fatal and 1038 injury crashes happened in urban areas. Table 2 shows the descriptive statistics for rural and urban segment lengths, which indicate that average mileage of rural segments (0.968 mile) is longer than urban segments (0.293 mile) because of a lower density of intersections in rural networks. Figure 1 displays the road segment crash rate distribution, calculated as the number of crashes per 100 million VMT, where the double line is the boundary between rural and urban areas. It shows that the percentage of segments with higher crash rates within the urban region is more than rural areas.

Table 3 displays the t -test statistics of rural and urban segment comparison for different types of facilities. It shows that there is a significant difference between rural and urban in terms of crash rates using both crash per lane*miles*year and crash per 100 million VMT in 2-lane segments. The crash rates in rural segments are consistently lower than urban segments. The 2-lane expressway is exceptional mainly because of the small sample size of 2-lane rural expressway. However, there is no statistical difference between rural and urban 4-lane arterial segments.

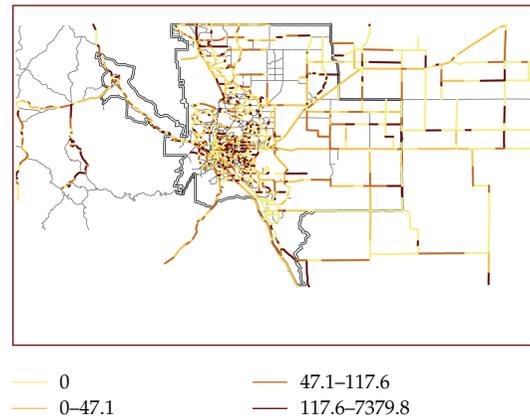


Figure 1: Road segment crash rate distribution in terms of the number of crashes per 100 million VMT.

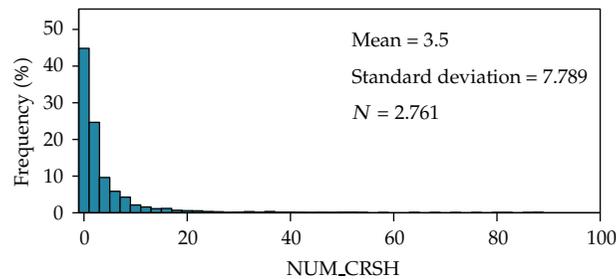


Figure 2: Crash frequencies distribution.

3.2. ZINB Regression Analyses

The crash frequencies distribution histogram (Figure 2) clearly illustrates that there are excessive zeros (over 40%) in the crash data. The P values in Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling normality tests are all less than 0.05. Therefore, it strongly supports the null hypothesis that the crash data do not follow the normal distribution. Therefore, the ZINB models are suitable to the crash count data regression analyses.

ZINB models were developed using the software SAS 9.2. We chose the crash frequency in segment (Num_crsh) as the dependent variable, and the regressors included segment length (length), number of lanes (Numberofla), thousand average annual day traffic (ADT_1000), free flow speed (FFS), and RoU (rural or urban). The segment type was not considered in this model since it was highly correlated with FFS and RoU.

Table 4 shows the parameter estimates of ZINB model for total crash frequency in segment, and only significant variables ($P < 0.05$) were included in the model. The ZINB model parameter estimates include 2 parts: NB regression and logistic regression. In the NB regression process, it can be found that the number of lanes, rural or urban, ADT, length, and FFS are all significantly correlated with the number of crashes. Further, the measure of Alpha in Table 4 is 1.435, with a P value less than 0.001, displaying a very strong overdispersion effect and indicating the superiority of the ZINB model over the zero-inflated Poisson (ZIP) model. ADT_1000 and LENGTH are positive associated with the crash frequency, suggesting that crash frequencies increase with increments of traffic volume and segment length. The

Table 3: *t*-test statistics for rural and urban segment comparison.

Road function	Lane	Variable	District	<i>N</i>	Mean	Standard deviation	Mean difference	<i>P</i> value		
Interstate	2	Per_lane_Mile	Rural	30	2.06	1.12	-2.456	0.000		
		Urban	86	4.52	4.63					
	4	Per_100_MVMT	Rural	30	45.31	25.07	-31.83	0.001		
		Urban	86	77.14	79.69					
		Per_lane_Mile	Rural	1	0.00	—				
		Urban	4	1.16	2.21					
Expressway	2	Per_100_MVMT	Rural	1	0.00	—	-30.184	—		
		Urban	4	30.18	56.91					
	4	Per_lane_Mile	Rural	2	0.44	0.62	-1.873	0.074		
		Urban	151	2.31	3.37					
		Per_100_MVMT	Rural	2	19.83	28.05				
		Urban	151	75.06	112.07					
P_Arterial	2	Per_lane_Mile	Rural	0	—	—	—	—		
		Urban	77	1.26	2.34					
	4	Per_100_MVMT	Rural	0	—	—	—	—		
		Urban	77	43.08	74.91					
		Per_lane_Mile	Rural	91	0.57	0.76			-0.621	0.000
		Urban	162	1.19	1.97					
Per_100_MVMT	Rural	91	34.01	44.34						
Urban	162	59.85	102.08							
M_Arterial	2	Per_lane_Mile	Rural	9	0.54	0.44	-0.363	0.546		
		Urban	592	0.90	1.79					
	4	Per_100_MVMT	Rural	9	23.99	19.57	-67.630	0.571		
		Urban	592	91.62	357.28					
		Per_lane_Mile	Rural	241	0.29	0.78			-0.307	0.000
		Urban	783	0.60	1.20					
Per_100_MVMT	Rural	241	108.70	424.56						
Urban	783	133.16	436.68							
Total comparison	2	Per_lane_Mile	Rural	0	—	—	—	—		
		Urban	532	0.59	1.41					
	4	Per_100_MVMT	Rural	0	—	—	—	—		
		Urban	532	105.56	246.18					
		Per_lane_Mile	Rural	374	0.51	0.93			-0.480	0.000
		Urban	2387	0.99	2.07					
Per_100_MVMT	Rural	374	82.64	343.16						
Urban	2387	102.96	331.91							

results are consistent with many previous research conclusions [7, 9, 27]. FFS is negatively associated with the crash frequency, indicating that crash frequencies are decreasing with increment of roadway free flow speed. Since FFS is correlated with the design standard of road facilities, it would be more appropriate to be explained that a better road facility with higher FFS has a lower crash rate compared to the facilities with lower FFS. In this study, FFS can be treated as a surrogate of speed limit but it can more accurately reflect the actual traffic operation status in road segments than speed limit. Previous research finding is less conclusive about the impact of speed limit on crash frequency [28]. In addition, four-lane roadways were found to be associated with a lower number of crashes than 2-lane roadways

Table 4: Parameter estimates of ZINB model for total crash frequency.

Parameter	DF	Estimate	Standard error	<i>t</i> value	<i>P</i> value
Negative binomial regression part					
Intercept	1	0.850	0.161	5.29	<0.0001
Numberofla (2 lane)	0	—	—	—	—
Numberofla (4 lane)	1	-0.438	0.107	-4.10	<0.0001
RoU (rural)	1	-0.270	0.095	-2.84	0.0045
RoU (urban)	0	—	—	—	—
ADT_1000	1	0.075	0.004	18.20	<0.0001
LENGTH	1	0.395	0.070	5.61	<0.0001
FFS	1	-0.011	0.004	-2.99	0.0028
Logistic regression part					
Inf_Intercept	1	8.214	0.913	8.99	<0.0001
Inf_ADT	1	-0.028	0.013	-2.20	0.0277
Inf_LENGTH	1	-71.713	8.739	-8.21	<0.0001
_Alpha	1	1.435	0.065	22.20	<0.0001

in this model. This is reasonable because this comparison was based on the assumption of same traffic exposure so that the segments with 4 lanes should have lower traffic volume per lane. More importantly, the urban regions appear to have a higher crash frequency than rural areas, which is consistent with the crash rate analyses results. The logistic regression part of the model predicts the likelihood of zero crash occurrences. The modeling results reveal that the variables of ADT_1000 and LENGTH are significant in estimating the probability of segments belonging to the zero crash occurrence group. According to the parameter coefficients estimated, the higher the traffic exposure (thousand of AADT and segment length), the lower the possibility of zero crash occurrences, which is consistent with all the previous study conclusions.

Furthermore, Table 5 shows the parameter estimates of ZINB model for injury and fatal crash frequency in a segment (Alpha is 1.074, with a *P* value less than 0.001). The NB regression indicates that Numberofla, RoU, ADT_1000, and LENGTH are significant variables to predict injury and fatal crash frequency, which displays a very similar result to that for total crash frequency except for FFS. It implies that although the better road facilities with higher FFS benefit from high standard design features resulting in a lower total crash frequency (as shown in Table 4), they would not mitigate the severe crash risk. A previous study reported that by controlling the other factors, purely increasing operation speed in road segments by 1% would approximately result in 2% increment in injury crash rate and 4% increment in fatal crash rate [29]. On the other hand, compared to the total crash frequency model, the logistic regression results for injury and fatal crash frequency model are quite different though the effect of LENGTH keeps similarity. First, the number of lanes is a significant variable for estimating the probability of zero injury and fatal crash occurrence in segment. Compared to 2-lane roads, the 4-lane roads have a lower severe crash frequency but have a lower probability of zero crashes. A possible explanation is that changing lane maneuver in 4-lane segments would increase the severe crash risk. Second, the effect of ADT_1000 in the Logistic regression of injury and fatal crash model is reverse from the total crash model. It shows that as traffic volume increases, the likelihood of zero severe crashes decreases. This interesting

Table 5: Parameter estimates of ZINB model for injury and fatal crash frequency.

Parameter	DF	Estimate	Standard error	<i>t</i> value	<i>P</i> value
Negative binomial regression part					
Intercept	1	-1.272	0.093	-13.73	<0.0001
Numberofla (2 lane)	0	—	—	—	—
Numberofla (4 lane)	1	-0.398	0.098	-4.08	<0.0001
RoU (rural)	1	-0.232	0.118	-1.96	0.0496
RoU (urban)	0	—	—	—	—
ADT_1000	1	0.060	0.004	15.58	<0.0001
LENGTH	1	0.338	0.068	5.00	<0.0001
Logistic regression part					
Inf_Intercept	1	5.404	0.850	6.35	<0.0001
Numberofla (2 lane)	0	—	—	—	—
Numberofla (4 lane)	1	-0.967	0.352	-2.75	0.0060
Inf_ADT_1000	1	0.035	0.016	2.21	0.0268
Inf_LENGTH	1	-38.325	6.436	-5.95	<0.0001
_Alpha	1	1.073	0.124	8.64	<0.0001

finding is consistent with the previous conclusion in a crash severity study, which explains that lower ADT could mean higher speeds that more often lead to severe/fatal crashes [30].

4. Conclusion and Discussions

There have been numerous studies to clarify the role of rural or urban settings in segment safety, but it was still controversial to make a conclusion. Before reaching the common agreement on the difference between rural and urban traffic safety, it is important to clarify the definition of “rural.” Generally, to distinguish from urban environments, rural areas have the attributes associated with demographic features (e.g., low population size and density, outside boundary of urban area), economic statuses (low economic indicators, farming, and agriculture), social structure (e.g., intimate, informal, and homogeneous forms of social interaction, limited social resources), cultural characteristics (e.g., traditional, conservative, provincial, slow to change), and so forth. The above features are often used to explain the statistical fact that the death rate from many common causes in US is significantly higher in rural compared to urban areas [1, 6], as well as in different countries [31–33].

However, these thresholds should not be universally applied to make local transportation safety analyses. For many developed regions, although districts are clearly separated into rural and urban regions according to their demographic, economic, or social attributes, the transportation facilities are well connected to each other and formed more standardized road networks. Thus, it was reported that there are relatively high numbers of crashes in urban regions because the heavy traffic volume and complex driving environments in urban lead to more conflicts between vehicles [34]. Therefore, for a specific safety evaluation project, this study supports the argument that more detailed crash risk comparisons between rural and urban transportation road segments should be performed at a comparable level. In this paper, the crash rate comparison and ZINB regression for both total crash frequency and injury and fatal crash frequency in road segment were conducted to discriminate between rural and urban traffic safety. It was found that compared to urban areas, the measures for

traffic safety in rural areas show lower crash rates, total crash frequencies, and injury and fatal crash frequencies. The results based on the ZINB regression models also showed the following.

- (i) Segment crash frequencies increase as traffic volume and segment length increase. However, higher traffic volume will lower the likelihood of severe crash occurrence.
- (ii) Compared to 2-lane roads, the 4-lane roads have a lower crash frequency but have a higher probability of severe crash occurrence.
- (iii) Better road facilities with higher free flow speed benefit from high standard design feature resulting in a lower total crash frequency but would not mitigate the severe crash risk.

Finally, it can be concluded that in the research area traffic safety of rural segments is better than urban segments, which implies that a priority for traffic safety improvement should be put on the urban highway segments.

Acknowledgments

The authors acknowledge that this study is supported by Chinese National 973 Project (2012CB725403), National Natural Science Foundation (71171014, 71210001), Ph.D. Programs Foundation of Ministry of Education of China (20110009110013), the State Key Laboratory of Rail Traffic Control and Safety (RCS2011ZT007), and Program for New Century Excellent Talents in University (NCET-11-0570).

References

- [1] M. S. Eberhardt, D. D. Ingram, and D. M. Makuc, *Urban and Rural Health Chart-Book. Health, United States, 2001*, National Center for Health Statistics, Hyattsville, Md, USA, 2001.
- [2] L. Kmet, P. Brasher, and C. Macarthur, "A small area study of motor vehicle crash fatalities in Alberta, Canada," *Accident Analysis and Prevention*, vol. 35, no. 2, pp. 177–182, 2003.
- [3] 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.
- [4] L. H. Brown, A. Khanna, and R. C. Hunt, "Rural versus urban motor vehicle crash death rates: 20 Years of FARS data," *Prehospital Emergency Care*, vol. 4, no. 1, pp. 7–13, 2000.
- [5] D. A. Borgianni, E. M. Hill, R. F. Maio, C. P. Compton, and M. A. Gregor, "Effects of alcohol on the geographic variation of driver fatalities in motor vehicle crashes," *Academic Emergency Medicine*, vol. 7, no. 1, pp. 7–13, 2000.
- [6] 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.
- [7] M. A. Abdel-Aty and A. E. Radwan, "Modeling traffic accident occurrence and involvement," *Accident Analysis and Prevention*, vol. 32, no. 5, pp. 633–642, 2000.
- [8] C. Zwerling, C. Peek-Asa, P. S. Whitten, S. W. Choi, N. L. Sprince, and M. P. Jones, "Fatal motor vehicle crashes in rural and urban areas: decomposing rates into contributing factors," *Injury Prevention*, vol. 11, no. 1, pp. 24–28, 2005.
- [9] X. Jiang, X. Yan, B. Huang, and S. H. Richards, "Influence of curbs on traffic crash frequency on high-speed roadways," *Traffic Injury Prevention*, vol. 12, no. 4, pp. 412–421, 2011.
- [10] M. Leff, L. Stallones, T. J. Keefe, R. Rosenblatt, and M. Reeds, "Comparison of urban and rural non-fatal injury: the results of a statewide survey," *Injury Prevention*, vol. 9, no. 4, pp. 332–337, 2003.
- [11] L. Aarts and I. van Schagen, "Driving speed and the risk of road crashes: a review," *Accident Analysis and Prevention*, vol. 38, no. 2, pp. 215–224, 2006.

- [12] C. Y. Chan, B. Huang, X. Yan, and S. Richards, "Effects of asphalt pavement conditions on traffic accidents in tennessee utilizing pavement management system," in *Proceedings of the 88th TRB Annual Meeting*, Washington, DC, USA, 2009.
- [13] P. J. Cooper, "The relationship between speeding behaviour (as measured by violation convictions) and crash involvement," *Journal of Safety Research*, vol. 28, no. 2, pp. 83–95, 1997.
- [14] N. J. Garber and A. A. Ehrhart, "Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways," *Transportation Research Record*, vol. 1717, pp. 76–83, 2000.
- [15] C. N. Kloeden, A. J. McLean, and G. Glonek, "Reanalysis of travelling speed and the rate of crash involvement in Adelaide South Australia," Tech. Rep. CR 207, Australian Transport Safety Bureau ATSB, Canberra, Australia, 2002.
- [16] J. Milton and F. Mannering, "The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies," *Transportation*, vol. 25, no. 2–4, pp. 395–413, 1998.
- [17] H. Renski, A. J. Khattak, and F. M. Council, "Effect of speed limit increases on crash injury severity: analysis of single-vehicle crashes on North Carolina Interstate highways," *Transportation Research Record*, no. 1665, pp. 100–108, 1999.
- [18] 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.
- [19] H. Huang and H. C. Chin, "Modeling road traffic crashes with zero-inflation and site-specific random effects," *Statistical Methods & Applications*, vol. 19, no. 3, pp. 445–462, 2010.
- [20] V. N. Shankar, F. L. Mannering, and W. Barfield, "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies," *Accident Analysis and Prevention*, vol. 27, no. 3, pp. 371–389, 1995.
- [21] A. Vogt and J. Bared, "Accident models for two-lane rural segments and intersections," *Transportation Research Record*, vol. 1635, pp. 18–29, 1998.
- [22] J. Carson and F. Mannering, "The effect of ice warning signs on ice-accident frequencies and severities," *Accident Analysis and Prevention*, vol. 33, no. 1, pp. 99–109, 2001.
- [23] 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.
- [24] D. Fletcher, "Geographic information systems for transportation: a look forward," in *Transportation in the New Millennium: State of the Art and Future Directions*, Transportation Research Board, CDROM, Washington, DC, USA, 2000.
- [25] X. Qiaojun, P. Fuquan, L. Jian, Z. Guoqiang, and W. C. Eidson, "GIS-based traffic safety management for highway intersections," in *Proceedings of the 1st International Symposium on Transportation and Development Innovative Best Practices (TDIBP '08)*, pp. 198–203, American Society of Civil Engineers, Beijing, China, April 2008.
- [26] S. Fang, Z. Yang, and Y. Chen, "Design of GIS-based road safety information management system," *Journal of Tongji University*, vol. 34, no. 5, pp. 629–633, 2006.
- [27] B. Persaud and L. Dzbik, "Accident prediction models for freeways," *Transportation Research Record*, vol. 1401, pp. 55–60, 1993.
- [28] B. N. Fildes and S. J. Lee, "The speed review: road environment, behavior, speed limits, enforcement and crashes," Tech. Rep. CR 127 (FORS) CR 3/93 (RSB), Federal Office of Road Safety and the Road Safety Bureau, Roads and Traffic Authority of New South Wales, New South Wales, Australia, 1993.
- [29] G. Nilsson, *Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety*, vol. 221 of *Bulletin*, Lund Institute of Technology, Lund, Sweden, 2004.
- [30] A. Das, M. Abdel-Aty, and A. Pande, "Using conditional inference forests to identify the factors affecting crash severity on arterial corridors," *Journal of Safety Research*, vol. 40, no. 4, pp. 317–327, 2009.
- [31] B. P. Y. Loo, W. S. Cheung, and S. Yao, "The rural-urban divide in road safety: the case of China," *The Open Transportation Journal*, vol. 5, pp. 9–20, 2011.
- [32] M. Y. Renwick, G. G. Olsen, and M. S. Tyrrell, "Road fatalities in rural New South Wales. Weighing the causes," *Medical Journal of Australia*, vol. 1, no. 7, pp. 291–294, 1982.
- [33] J. P. Thouez, M. F. Joly, A. Rannou, Y. Bussiere, and R. Bourbeau, "Geographical variations of motor-vehicle injuries in Quebec, 1983–1988," *Social Science and Medicine*, vol. 33, no. 4, pp. 415–421, 1991.
- [34] S. Eiksund, "A geographical perspective on driving attitudes and behaviour among young adults in urban and rural Norway," *Safety Science*, vol. 47, no. 4, pp. 529–536, 2009.

Research Article

Driver Cognitive Distraction Detection Using Driving Performance Measures

Lisheng Jin, Qingning Niu, Haijing Hou, Huacai Xian, Yali Wang, and Dongdong Shi

Transportation College, Jilin University, Changchun, Jilin 130022, China

Correspondence should be addressed to Haijing Hou, houhj10@mails.jlu.edu.cn

Received 17 August 2012; Revised 26 October 2012; Accepted 27 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Lisheng Jin 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.

Driver cognitive distraction is a hazard state, which can easily lead to traffic accidents. This study focuses on detecting the driver cognitive distraction state based on driving performance measures. Characteristic parameters could be directly extracted from Controller Area Network-(CAN-)Bus data, without depending on other sensors, which improves real-time and robustness performance. Three cognitive distraction states (no cognitive distraction, low cognitive distraction, and high cognitive distraction) were defined using different secondary tasks. NLModel, NHModel, LHModel, and NLHModel were developed using SVMs according to different states. The developed system shows promising results, which can correctly classify the driver's states in approximately 74%. Although the sensitivity for these models is low, it is acceptable because in this situation the driver could control the car sufficiently. Thus, driving performance measures could be used alone to detect driver cognitive state.

1. Introduction

Driver distraction is a major factor in traffic accidents, and it is estimated that up to 23 percent of crashes and near-crashes are caused by driver distraction. As the use of in-vehicle information systems (IVISs) such as cell phones, navigation systems, and satellite radios, will increase these figures will likely increase [1–6]. Thus enabling drivers to benefit from IVISs without diminishing safety is an important challenge [7]. One way to solve this problem is to detect the driver state in real time, and when distraction occurs, the corresponding warning system works to mitigate the effects of distraction [8].

Obviously, measuring driver state in real time is the core function in such systems. There has been an explosion research on these topics including the definition, classification, and detection of distraction. Donmez et al. [8] proposed a general definition that is “driver

distraction is a diversion of attention away from activities critical for safe driving toward a competing activity.”

Generally, visual distraction and cognitive distraction are the two major types. Visual distraction can be described as “eye-off-road” and cognitive distraction as “mind-off-road” [9]. Both of them can undermine drivers’ performance.

Visual distraction is straightforward, occurring when drivers look away from the roadway (e.g., to adjust a radio), which can be reasonably measured by the length and frequency of glances away from the road [10]. Unlike visual distraction, cognitive distraction occurs when drivers think about something not directly related to the current vehicle control task (e.g., conversing on a hands-free cell phone or route planning) [6]. Therefore, in this paper we only detect driver cognitive distraction.

2. Measures Selection

There are five types of measures for driver inattention detection [11]: (1) subjective report measures (e.g., SSS, KSS); (2) driver biological measures (e.g., EEG, ECG); (3) driver physical measures (e.g., PERCLOS, gaze direction); (4) driving performance measures (e.g., steering wheel angle, yaw angle); (5) hybrid measures.

Since cognitive distraction needs to be done in real time and nonintrusively, the subjective report measures and driver biological measures are not suitable for a real-life context.

2.1. Driver Physical Measures

The most common used driver physical data for driver cognitive distraction are eye movements [6, 12]. Azman et al. [13] found that mouth and eyes are correlated to each other when a person is thinking or cognitively distracted and they could be used to detect driver’s cognitive distraction. Victor et al. [9] found that cognitive distraction causes drivers to concentrate their gaze in the center of the driving scene, as defined by the horizontal and vertical standard deviation of gaze distribution, and diminishes drivers’ ability to detect targets across the entire driving scene. Fletcher and Zelinsky [14] utilized faceLAB to obtain information such as eye gaze direction, eye closure, and blink detection, as well as head position.

2.2. Driving Performance Measures

A change in the mental state can induce the change in driving performance. Many studies prove the fact that compared to the attentive drivers the distracted ones steer their car in a different way; the same applies for throttle use and speed [15]. Some lines of evidence show that drivers adjust their behavior according to cognitive demand of secondary tasks. Drivers tend to increase the distance to the leading vehicle in the car-following scenario when they engage in cognitively demanding secondary tasks [16–18]. This suggests that drivers may compensate for the impairments that secondary tasks have imposed. One study found that drivers drove faster than the normal when distracted by a cognitive task [19]. Liang et al. [6] found that the driving measures improved the cognitive distraction detection model performance dramatically and built SVM models used driving performance only. The driving

measures consist of standard deviation of steering wheel position, mean of steering error, and standard deviation of lane position. But compared with gaze behavior, they found that gaze-related features led to much better prediction accuracy than using the driving performance measures alone. The similar conclusion was found in [4]. Wollmer et al. [1] introduced a technique for online driver distraction detection that used LSTM recurrent neural nets to continuously predict the driver's state based on driving and head-tracking data. The measured signals include steering wheel angle, throttle position, speed, heading angle, lateral deviation, and head rotation. These links between driving performance and cognitive state show that driving performance measures are good candidates to predict cognitive distraction.

2.3. Hybrid Measures

In [6], driver physical measures and driving performance measures were combined to detect driver distraction in real time. Comparing support vector machines (SVMs) to traditional logistic regression models, the results showed that the SVMs models performed better. In [20], machine-learning techniques were used to detect driver cognitive distraction based on the standard deviations of eye gaze, head orientation, pupil diameter, and average heart rate (RRI). The eye and head parameters were obtained using faceLAB, whereas the RRI data came from ECG. Sathyanarayana et al. [21] detected distraction by combining motion signals from the leg and head with driving performance signals using a k -nearest neighbor classifier, the driving performance signals adopted including vehicle speed, braking, acceleration, and steering angle.

Among all of these measures, eye movements are one of the most promising ways to assess driver distraction [4, 6, 12]. While most of the eye movements parameters were obtained by faceLAB or SmartEye, these systems are not common in vehicles today, owing to their higher price for installation into a vehicle. At the same time there are limits in the process of extracting eye movements' parameters [6].

- (1) Complex calibration: before each experimental drive, the calibration of the gaze vector with the simulator screen must be verified. After that, in the process of the experiment eye tracker must be calibrated to every participant and the calibration takes 5 to 15 min. After the complex calibration, the tracking error was approximately 5% of visual angle for most participants.
- (2) Driver restriction: the participants cannot wear glasses or eye make-up because these conditions can negatively affect tracking accuracy.
- (3) Environmental restriction: eye trackers may lose tracking accuracy when vehicles are traveling on rough roads or the lighting conditions are variable.
- (4) Time delay: the Seeing Machines' faceLAB eye tracking system takes approximately 2.6 s to transfer camera image to numerical data.

These requirements limit the application of cognitive distraction system using eye movements parameters obtained from faceLAB or SmartEye; therefore, up till now, this scheme is only for research offline. More robust and real-time eye tracking techniques are needed to make these detection systems become a reality. While driving performance parameters could be obtained in real time from CAN-Bus directly, driving performance measures are used in this study for cognitive distraction detection. In this method, the characteristic parameters could be directly extracted without depending on other sensors, and system real-time performance and robustness are improved.

3. Model Development

Driving performance data were collected in a simulator experiment. The driving simulation directly outputs driving performance original data, which is collected each 10 Hz. After extracting characteristic parameters from the original data, SVM model was trained for each participant. Twelve subjects participated in the experiment to detect driver's cognitive distraction state.

3.1. Experiment

3.1.1. Participants

Twelve participants (4 women and 8 men) aged 21–40 years old took part in this study. All participants were experienced drivers with valid licenses. Participants were recruited via an advertisement in school website.

3.1.2. Driving Simulator

The driving simulator used in this experiment is shown in Figure 1. The highway scenario is a 133 km long highway of a sampled actual ChangPing highway located between ChangChun and SiPing city, with two lanes in one direction. The traffic situation selected in this experiment was only sparse oncoming traffic and no traffic driving in the same direction as the test subject.

3.1.3. Driving Task

The driving experiment was to drive the simulator at 80–120 km/h. Every participant was asked to drive 4 sessions. In the first session, the participant drove 20 minutes to be familiar with the driving condition. In the following three sessions, the participants were asked to perform secondary tasks, including no workload tasks, low workload tasks, and high workload tasks. Each session took 25–35 minutes.

In the no workload task session, subjects drove the simulator without secondary task introduced, called no cognitive distraction (NCD). During the events, researcher randomly recorded ten different 60-second periods driving performance original data as NCD data.

In the low workload tasks session, subjects were asked to talk with the researcher. During the events, researcher ensures the talk diverts part of subjects' attention away from activities critical for safe driving. The talk added subjects' workload and made subjects think about something not directly related to the current vehicle control task but not led to be lost in thought, called low cognitive distraction (LCD). During the talking process, another researcher randomly recorded ten different 60-second periods driving performance original data as the LCD data.

In the high workload tasks session, subjects were asked to answer the researcher's questions (intelligent test questions). These intelligent test questions diverted subjects' attention away from activities critical for safe driving, and led subjects to think and be lost in thought, called high cognitive distraction (HCD). During the events, when another researcher ensured subjects thought about a question, driving performance original data were



Figure 1: Driving simulator.

recorded as the HCD data. In this session, the original data record period depends on driver's distraction state. This session took much more time until enough original data were recorded.

3.1.4. Original Data

The driving performance original data were directly obtained from the driving simulator including (1) vehicle velocity, (2) vehicle acceleration, (3) steering wheel angle, (4) steering wheel angular velocity, (5) throttle position, (6) yaw angle, and (7) yaw angular velocity.

3.2. SVM Model

3.2.1. Support Vector Machines (SVMs)

Support vector machine (SVM) is a popular machine learning method for classification, regression, and other learning tasks, which is first proposed by Vapnik [22].

The basic idea of classification using SVMs in 2D space is shown in Figure 2. Labeled binary-class training data $D = \{(x_i, y_i)\}_{i=1}^l$, where x_i is a vector containing multiple features and y_i is a class indicator with value either -1 or 1 , are illustrated as circles and dots in Figure 2, respectively.

They are mapped onto a high-dimensional feature space via a function Φ . When the mapped data are linearly separable in the feature space, a hyperplane maximizing the margin from it to the closest data points of each class exists. The hyperplane yields a nonlinear boundary in the input space. The maximum margin represents the minimized upper bound of generalization error. The function is written in the form of a kernel function $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ used in the SVM calculation. When data are not linearly separable in the feature space, the positive penalty parameter C allows for training error ϵ by specifying the cost of misclassified training instances.

As cognitive distraction affects driving performance complexly, the learning technique of the SVM method makes it very suitable for measuring the cognitive state of humans.

3.2.2. Characteristic Parameters

The original data were preprocessed and generated a vector of characteristic parameters as listed in Table 1 as model input.

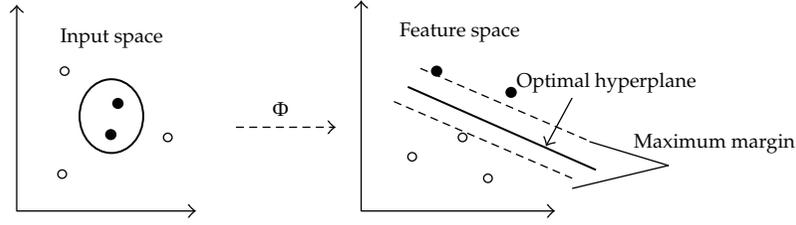


Figure 2: SVM algorithm [6].

Table 1: Characteristic parameters.

Feature number	Features	Meaning
1	VVM	Mean vehicle velocity
2	VVD	Standard deviation of vehicle velocity
3	VAM	Mean vehicle acceleration
4	VAM	Standard deviation of vehicle acceleration
5	SAM	Mean steering wheel angle
6	SAD	Standard deviation of steering wheel angle
7	SVM	Mean steering wheel angular velocity
8	SVD	Standard deviation of steering wheel angular velocity
9	TPM	Mean throttle position
10	TPD	Standard deviation of throttle position
11	YAM	Mean yaw angle
12	YAD	Standard deviation of yaw angle
13	YVM	Mean yaw angular velocity
14	YVD	Standard deviation of yaw angular velocity

3.2.3. Window Size

Window size denotes the period over which characteristic parameters were averaged. In order to improve the real-time performance, the window size 1 s was chosen in this paper. The characteristic parameters were summarized across the window size to form instances as model input. For every participant there were 600 training instances to each distraction state (NCD, LCD, and HCD).

3.2.4. Model Training

Cognitive distraction affects driver behavior in a subtle, inconsistent manner, which can be easily washed out by individual differences associated with driving style [23]. Thus SVM model was trained for each participant. We randomly selected 200 training instances to each distraction state (NCD, LCD, and HCD) and used the remaining instances, which accounted for at least two thirds of total instances for testing.

The “LIBSVM” Matlab toolbox [24] was used to train and test SVM models, and LIBSVM is currently one of the most widely used SVM software. Linear, polynomial, radial basis function (RBF), and sigmoid are the four basic kernels. The RBF can nonlinearly map samples into a higher dimensional space, which can handle the case when the

relation between class labels and attributes is nonlinear. At the same time, the RBF can reduce numerical difficulties and tend to obtain more robust results than other kernels, such as polynomial. Furthermore, compared to the polynomial kernel, the RBF has less hyperparameters which influence the complexity of model selection [25]. Therefore the RBF was chosen as the kernel function for the SVM models:

$$K(x_i, x_j) = e^{-\gamma|x_i - x_j|^2}, \quad (3.1)$$

where x_i and x_j represent two data points and γ is a predefined positive parameter. There are two parameters for the RBF kernel C (the penalty parameter) and γ , and it is not known beforehand. In order to improve the model prediction performance, grid-search is recommended to identify good (C, γ) using cross-validation (CV). LIBSVM provides a tool to check a grid of parameters, and CV accuracy was obtained for each parameter setting. When the highest CV accuracy returns, (C, γ) are selected to the model.

3.2.5. Model Performance Measures

Model performance was evaluated with three different measures: accuracy, sensitivity, and specificity [26], which were calculated according to

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TU}}{\text{TON}} \times 100, \\ \text{Sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100, \\ \text{Specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100. \end{aligned} \quad (3.2)$$

The definition of sensitivity and specificity is shown in Table 2. where TU represents the number of true prediction instances; TON represents the total number of instances; TP represents true positive, which is defined as when the driver is distracted and the model detect result is distracted; TN represents true negative, which is defined as when the driver is not distracted and the model detect result is no distracted; FP represents false positive, which is defined as when the driver is not distracted and the model detect result is distracted; FN represents false negative, which is defined as when the driver is distracted and the model detect result is not distracted.

4. Experimental Results

Different SVM models were developed by using the same training instances from two of the three distraction states (NCD, LCD, and HCD) to compare the accuracy of this detection system based on driving performance measures when the driver was in different cognitive distraction states. NLModel, NHModel, LHModel, and NLHModel were developed, where NLModel was developed to distinguish LCD from NCD, NHModel was developed to distinguish HCD from NCD, and LHModel was developed to distinguish LCD from HCD.

Table 2: Definition of sensitivity and specificity.

Model detect result	Driver actual state	
	Distracted	No distracted
Distracted	TP (true positive)	FP (false positive)
No distracted	FN (false negative)	TN (true negative)

The selection process for the best (C, γ) of a LHModel is shown in Figure 3. We searched for C and γ in the growing sequences ranging from 2^{-8} to 2^8 according to (4.1). At last the best $C = 0.0625$, $\gamma = 0.32988$ were selected when the CV accuracy was 95.75%:

$$\begin{aligned} C_i &= 2^{-8+i}, \\ \gamma_j &= 2^{-8+j}, \end{aligned} \quad (4.1)$$

subject to : $0 \leq i \leq 16, \quad 0 \leq j \leq 16$.

After all of the models developed, the mean detecting performance is shown in Figure 4. The mean accuracy for all NLModels was 78% (std = 0.195), the mean sensitivity was 56% (std = 0.39), and the mean specificity was 100% (std = 0). The mean accuracy for all LHModels was 66.14% (std = 0.1616), the mean sensitivity was 84.86% (std = 0.2024), and the mean specificity was 47.43% (std = 0.3748). The mean accuracy for all NHModels was 76.87% (std = 0.1568), the mean sensitivity was 54.07% (std = 0.3145), and the mean specificity was 99.75% (std = 0.0071).

The results thus far have demonstrated that driver cognitive distraction affects driving performance obviously, and driving performance measures could be used alone to detect driver state. The accuracy for LHModel is 66.14%. It shows that when driver workload changes from LCD to HCD, the driver performance changes. However, the accuracy for NLModel (78%) and NHModel (76.87%) is higher than LHModel (66.14%), which means the change rate of driving performance between NCD and LCD (or NCD and HCD) is higher than it is between LCD and HCD.

The sensitivity for NLModel (LHModel, NHModel) means that the accuracy rate of it predicts NCD (LCD, NCD) test instances as NCD (LCD, NCD), and the specificity for NLModel (LHModel, NHModel) means that the accuracy rate of it predicts LCD (HCD, HCD) test instances as LCD (HCD, HCD). From the detecting performance, specificity for NLModel is apparently higher than sensitivity, NLModel can predict NCD test instances accurately as NCD, while it classifies some LCD test instances as NCD. It suggests that the characteristics of NCD are more significant than that of LCD.

One low NLModel prediction performance is shown in Figure 5, where NCD label is 1, and LCD label is 2. It shows that NLModel can predict NCD class accurately (100%) and the LCD class prediction accuracy is only 46%. Though NLModel classifies some LCD as NCD, it can predict NCD accurately. False alarm rate for this system is low, which increases the system acceptance.

It appears that low sensitivity of this system increases the missing alarm rate. However cognitive distraction state (LCD or HCD) is defined as the driver performs a secondary task in this paper, and the subject's distraction state is distinguished by researcher's subjective judgment. At the same time, driving performance changes when the driver is in distraction state, but which is not totally different from NCD at any moment during LCD or HCD

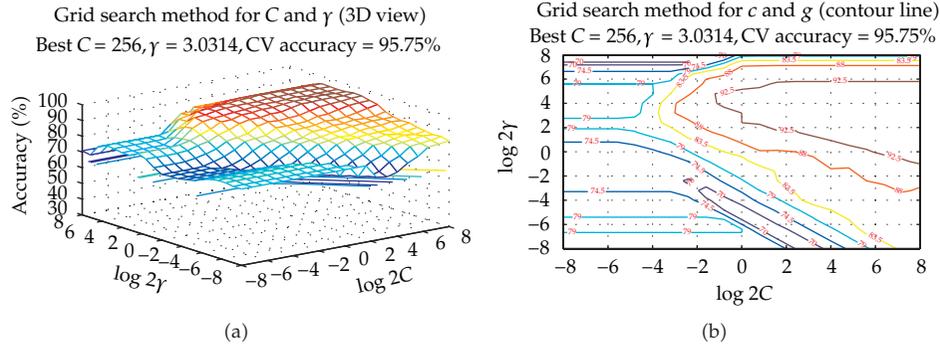


Figure 3: Parameter selection for NLModel.

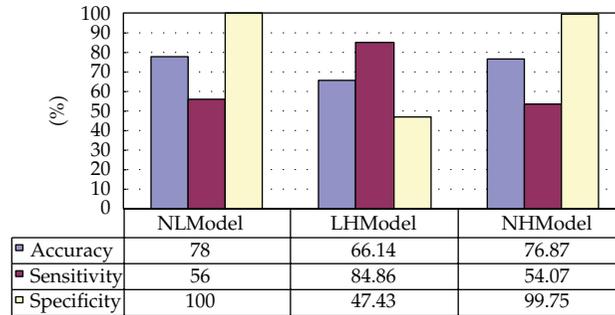


Figure 4: The detecting performance for NLModel, LHModel, and NHModel.

period. Especially for experienced drivers, the characteristic change between NCD and LCD (or HCD) is short and smooth. Thus part of the train and test instances from distraction state (LCD or HCD) is basically similar with NCD, and the sensitivity is relatively low. Fortunately, as long as driving performance is not affected by cognitive distraction, driver can control the activities critical for safe driving effectively, a situation where the driver is safe. Therefore low sensitivity is acceptable.

NLHModel is developed to classify NCD, LCD, and HCD, and NLHModel performance shows in Figure 6. The mean accuracy was 73.86% (std = 0.1633), NCD prediction accuracy was 97.86% (std = 0.0411), LCD prediction accuracy was 65% (std = 0.3718), HCD prediction accuracy was 60.14% (std = 0.3457), sensitivity was 61.86% (std = 0.237), and specificity was 97.86% (std = 0.0411). The sensitivity for NLHModel means the accuracy rate of that it predicts NCD test instances as NCD, and the specificity for NLHModel means the accuracy rate of it predicts LCD and HCD test instances as LCD and HCD. Thus, NLHModel could be used to predict different cognitive state, and it has the similar conclusion as NLModel and NHModel.

5. Discussions and Conclusion

Compared with driver physical measures, using driving performance measures to detect driver cognitive distraction is more effective, simple, and of real time, so it is used to detect driver state in this paper. NCD, LCD, and HCD were defined as three different cognitive

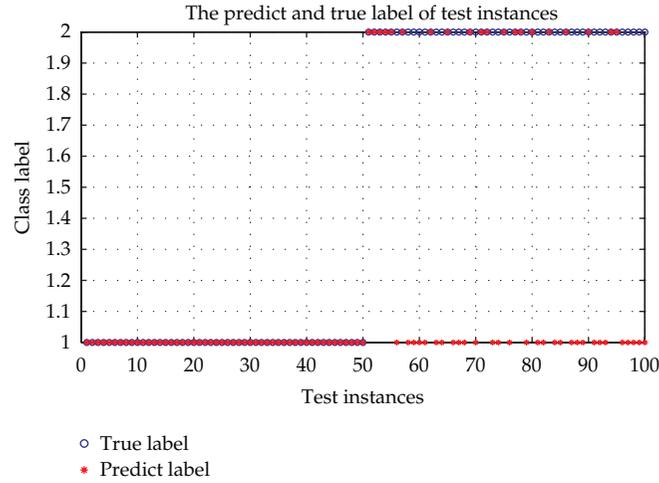


Figure 5: NLModel prediction result.

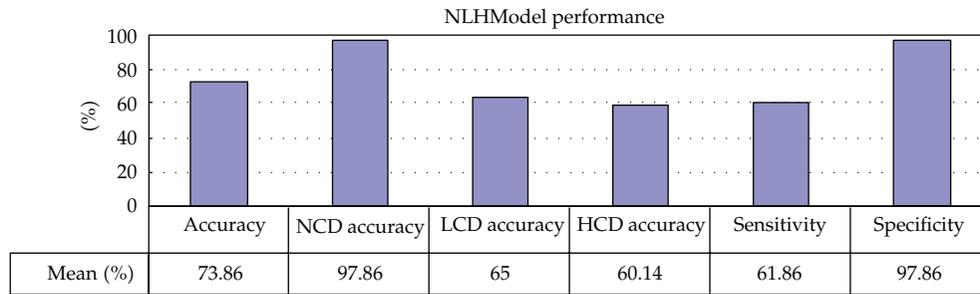


Figure 6: NLHModel predict performance.

distraction states using different secondary tasks. Twelve drivers were recruited to take part in the experiment. For every participant, 7 original data about driving performance were obtained from the driving simulator directly, and 14 characteristic parameters were extracted as SVM models input. In order to improve real-time performance of the developed models, window size used in this research was 1 s. At last, different SVM models (NLModel, NHModel, LHModel, and NLHModel) were developed by using the same training instances from two of the three distraction states (NCD, LCD, and HCD) to compare the accuracy of this detection system when the driver was in different cognitive distraction states. The mean accuracy of each SVM model is approximately 74%; thus driving performance can be used alone to detect driver cognitive state. The specificity is up to 99%, and false alarm rate for this system is low, which increases the system acceptance. The sensitivity of each SVM model is low, which is acceptable, because in this situation the driver could control the car sufficiently.

At the same time, the participant's cognitive state is distinguished by researcher's subjective judgment, which affects the model's accuracy seriously. Therefore it could be helpful to use driver biological measures as the standard for future research to select correct and reasonable training instances as model's input. Furthermore, SVM models were trained for each participant, and it might be interesting to select characteristic parameters which cannot be affected by individual differences driving style.

Acknowledgments

This research was supported partly by Doctoral Study Special Research Foundation of Higher Education (no. 20110061110036), Major Projects of Jilin Science and Technology Department (no. 20116017), New Century Excellent Talent Foundation Program under grant and National Nature Science Foundation of China (no. 50908098).

References

- [1] M. Wollmer, C. Blaschke, T. Schindl et al., "Online driver distraction detection using long short-term memory," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 574–582, 2011.
- [2] H. Hou, L. Jin, Q. Niu, Y. Sun, and M. Lu, "Driver intention recognition method using continuous Hidden Markov model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 386–393, 2011.
- [3] L. S. Jin, Q. N. Niu, H. J. Hou, S. X. Hu, and F. R. Wang, "Study on vehicle front pedestrian detection based on 3D laser scanner," in *Proceedings of the International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE '11)*, pp. 735–738, Changchun, China, December 2011.
- [4] Y. Zhang, Y. OwecHko, and J. Zhang, "Driver cognitive workload estimation: a data-driven perspective," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (ITSC '04)*, pp. 642–647, Washington, DC, USA, October 2004.
- [5] M. A. Regan, C. Hallett, and C. P. Gordon, "Driver distraction and driver inattention: definition, relationship and taxonomy," *Accident Analysis and Prevention*, vol. 43, no. 5, pp. 1771–1781, 2011.
- [6] Y. Liang, M. L. Reyes, and J. D. Lee, "Real-time detection of driver cognitive distraction using support vector machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 340–350, 2007.
- [7] 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.
- [8] B. Donmez, L. Boyle, and J. D. Lee, "Taxonomy of mitigation strategies for driver distraction," in *Proceedings of the 47th Annual Meeting on Human Factors and Ergonomics Society*, pp. 1865–1869, Denver, Colo, USA, 2003.
- [9] T. W. Victor, *Keeping eye and mind on the road [doctoral thesis]*, Uppsala Universitet, Interfaculty Units, Acta Universitatis Upsaliensis, Uppsala, Sweden, 2005.
- [10] D. Haigney and S. J. Westerman, "Mobile (cellular) phone use and driving: a critical review of research methodology," *Ergonomics*, vol. 44, no. 2, pp. 132–143, 2001.
- [11] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: a review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 596–614, 2011.
- [12] M. M. Hayhoe, "Advances in relating eye movements and cognition," *Infancy*, vol. 6, no. 2, pp. 267–274, 2004.
- [13] A. Azman, Q. Meng, and E. Edirisinghe, "Non intrusive physiological measurement for driver cognitive distraction detection: eye and mouth movements," in *Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE '10)*, vol. 3, pp. V3595–V3599, August 2010.
- [14] L. Fletcher and A. Zelinsky, "Driver state monitoring to mitigate distraction," in *Proceedings of the Internal Conference on the Distractions in Driving*, pp. 487–523, Sydney, Australia, 2007.
- [15] K. Young, M. Regan, and M. Hammer, *Driver Distraction: A Review of Literature*, Accident Research Centre—Monash University, Victoria, Australia, 2003.
- [16] W. J. Horrey and D. J. Simons, "Examining cognitive interference and adaptive safety behaviours in tactical vehicle control," *Ergonomics*, vol. 50, no. 8, pp. 1340–1350, 2007.
- [17] D. L. Strayer and F. A. Drews, "Profiles in driver distraction: effects of cell phone conversations on younger and older drivers," *Human Factors*, vol. 46, no. 4, pp. 640–649, 2004.
- [18] 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.
- [19] Y. C. Liu, "Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveler information systems," *Ergonomics*, vol. 44, no. 4, pp. 425–442, 2001.
 - [20] M. Miyaji, H. Kawanaka, and K. Oguri, "Driver's cognitive distraction detection using physiological features by the AdaBoost," in *Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems (ITSC '09)*, pp. 90–95, St. Louis, Miss, USA, October 2009.
 - [21] A. Sathyanarayana, S. Nageswaren, H. Ghasemzadeh, R. Jafari, and J. H. L. Hansen, "Body sensor networks for driver distraction identification," in *Proceedings of the IEEE International Conference on Vehicular Electronics and Safety (ICVES '08)*, pp. 120–125, September 2008.
 - [22] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
 - [23] Y. L. Liang, *Detecting driver distraction [doctoral thesis]*, University of Iowa, 2009.
 - [24] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," <http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html#GUI>.
 - [25] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
 - [26] J. Berglund, *In-vehicle prediction of truck driver sleepiness-steering related variables [M.S. thesis]*, University of Linköping, 2007.

Research Article

Transportation Structure Analysis Using SD-MOP in World Modern Garden City: A Case Study in China

Jiuping Xu,^{1,2} Jing Yang,^{1,2} and Liming Yao^{1,2}

¹ *Low-Carbon Technology and Economy Research Center, Sichuan University, Chengdu 610064, China*

² *Uncertainty Decision-Making Laboratory, Sichuan University, Chengdu 610064, China*

Correspondence should be addressed to Jiuping Xu, xujiuping@scu.edu.cn

Received 6 July 2012; Revised 7 September 2012; Accepted 26 September 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Jiuping Xu 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 idea of the “garden city” was developed theoretically to offer solutions to serious city development problems such as traffic congestion, population, and environmental pollution, among which the transportation is considered the most important. The question is how to develop balanced transportation in a garden city. Transportation is a complex system, particularly in a garden city. Therefore, we establish a new approach named the transportation multiobjective optimization system dynamics (SD-MOP) model, which firstly calculates the optimal proportion of different transport means with an MOP approach and then applies them to the dynamic transportation system to analyze the results and analyze the influence on the whole system using different transportation means variation. In this paper, we take Chengdu as an example, one of the few cities in the world declared as building a garden city, and then develop some recommendations about world modern garden city transportation system development.

1. Introduction

It is generally recognized that cities are experiencing huge change in terms of their development and mobility patterns, while transportation, and will continue to plays, a critical role in city development [1–3]. Energy consumption is one of the most severe transportation problems. IEA [4] argues that transport plays an important role in addressing the challenges of climate change mitigation as it consumes nearly half of global oil and contributes 25% of total fossil fuel combustion-related CO₂ emissions of the world, and road transport is responsible for about 75% of the emissions from the transport sector. Petri et al. [5] compare the development of transport and energy use with a focus on CO₂ emissions and suggest a more sustainable passenger transport system. Dominic [6] examines recent

temporal and spatial trends and forecasts in energy consumption, energy efficiency, and energy costs in the transport sector across Europe. Meanwhile, land use, health effects, employment, population growth, and transport alternatives are all considered as related to the transportation problems. Frank [7] focused on land use, noting that, with different land uses, traffic designs need to be different. Messenger and Ewing [8] think that employment, the balance of living space, ownership, and the public transport service level affect people's transport choice. Martin [9] investigates the association between means of transportation to work and overweight and obesity. In this paper, transportation structure is our main concern to the research. Transportation structure is the proportion of traffic amount carried by different transport means in extent of time and space. As the transportation structure directly influences resource allocation, a reasonable urban transportation structure can contribute to the rational use of resources and ensure a well-functioning system [10]. Although these studies have contributed a lot to transportation, we feel that all of the studies had not analyzed the transportation in a systematic and dynamic way. Thus, this paper seeks to further research in solving transportation problem and differs from its predecessors and we hope to introduce completely and accurately new viewpoints and models and research.

China is the largest developing country in the world. With rapid process of industrialization and urbanization, China has maintained an extensive growth in economic development while the deficiency of transportation began to emerge and became an urgent problem for us to deal with. Traffic congestion exists widely in metropolitan [11] capacity excess or overload caused by road passenger volume [12], which has already led to problems such as environmental pollution, lack of rational planning [13], economic intervention, and greenhouse gas emission excess [14]. Steps need to be taken to prevent the situation from deteriorating, otherwise, in return, this may hinder the development of the world modern garden city.

The "garden city" was first proposed by Howard [15] in the late 19th century, which came into being with the overcrowded, pollution, and epidemic spreading problems. It focuses on the coalition of city and countryside in essence and, later, makes some city planning about city scale, layout and structure, population density, and green belt [16]. A garden city is designed for health, life, and industry; it contains both rural and urban areas and has a strictly controlled city scale. It is the farmlands and towns around the central cities that control the expanding of urban land without limit. The garden city can ensure every resident to be close to nature and surrounded by self-sufficient farmland; especially in an ideal garden city, the land belongs to the public and under the responsibility of a professional committee. Therefore, the world modern garden city has its own features that differ from the ordinary city. Firstly, the form and pattern of garden city are multicenter, networking, and clustered in development along with being humanized in urban spatial structure. Secondly, harmonious nature and society: there are two kinds of harmony which refer to the strengthening of ecology and environment, social welfare, and wellbeing. Thirdly, the development path: the city aims to modern service industry and headquarters economy as the core, for the direction of high-tech industry, based on the powerful modern manufacturing industry and agriculture, all of which projects to be an internationally regional hub and central city on the basis of to be western and national central city. Fourthly, land use layout and transportation: it is the decentralized layout that is put into use in garden city, while in ordinary city, the public buildings are always arranged in concentrated form.

To realize the construction of world modern garden city, transportation should play its due role in it and act as a stimulus to promote it. As for the transportation in garden city, we think it is the traffic arteries that connect the central city with peripheral group

city, with the agricultural land scattered around it, which finally realize the garden city. And the key way to build a world modern garden city is to promote the modernized and intellectualized transportation, that is, to make the linking of the urban and rural areas come true. In garden city, environmental and faster high-speed railway is the best choice to create the traffic circle in connecting between cities. In order to construct garden city, the transportation should match the development of garden city, and, in turn, the garden city will surely promote the transportation construction. Therefore, a strategy is needed to balance transportation system development and garden city construction, as transportation is an essential element of its success. Since regional transportation system is constantly changing, it is necessary to find a dynamic simulation method. System dynamics (SD) approaches as a modeling tool to provide a flexible way of building simulation models from causal loop or stock and flow diagrams. Therefore, to reflect the dynamic characteristics of garden city transportation system, the SD approach is the main methodology used in this paper, combined with multiobjective optimization (MOP) for its effectiveness.

The aim of this paper is to propose a system dynamics and multiobjective programming integrated support model to predict and adjust transport structure for the modern garden city in the world. The remainder of this paper is structured as follows. Section 2 describes the general system and solution approach problem. Section 3 develops a detailed garden city transportation system based on the SD-MOP model. In Section 4, Chengdu in China is discussed as a case study. Finally, we present some conclusions and proposals for the development of the transportation system in Chengdu and other garden cities in the world.

2. Problem Description

In this Section, a description of the problem is discussed, then a general framework to address the given problem is proposed. We give a basic background for our study.

2.1. System Description

It is of great significance to analyze logical urban transportation system in Chengdu, because it can assist in the development and management of the transportation plan and has a practical significance in helping relieving city traffic congestion [17]. Transport structure is an important factor in the whole system; a reasonable logic transport structure is a part of city planning and the adjustment of industrial structure, meanwhile, it guarantees minimum time waste, costs, and environment pollution.

As has been mentioned, an urban transportation system is a complex system and is especially important in the development of the garden city. With population, transport means, transport congestion, transport demand, and vehicle travel time are emerging as concerns in transportation system analysis. These elements are highly interrelated, but they are not the only factors that affect the system, there is also social, economic, political, environmental, and technical factors [17]. From previous research [18, 19], we assume that the garden city transportation system consists of five subsystems: the economic subsystem, the environmental subsystem, the traffic congestion subsystem, the policy management subsystem, and the traffic mode subsystem. The whole system is constantly changing and has an interrelationship with each other. Figure 1 shows the relationships between them. With economic development, there are more travel demand and transport choices or modes, and if

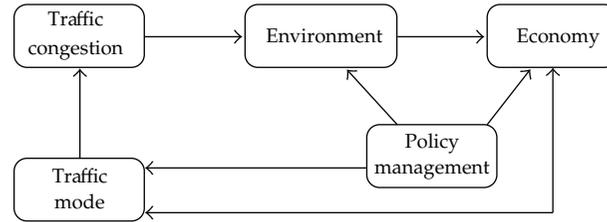


Figure 1: The subsystems of garden city transportation system.

not managed properly, they will lead to traffic congestion, which can result in environmental pollution, and, in turn, impacts the economy. However, through manual intervention, policy management can be used to control these effects when necessary.

2.2. Method Design

System dynamics (SD) is a simulation technology that studies complex systems based on feedback control theory. It establishes synthetical models using system structures, the relation of consequent to antecedent and feedback loops, and, further, to find the solution to system performance using simulations. SD has been applied to a number of studies, not only the social sciences field, but also the agricultural practices [20], environmental issues [21], and economic controls [22] and has proven to be especially appropriate for modeling problems. Meanwhile, a number of system dynamics (SD) approaches have been used to do transportation modeling [23, 24], which give us successful examples for our research. SD can be used to forecast the trends in the next ten years by using certain parameters, but cannot be used to estimate exact levels reliably [25–27]. Therefore, while a system dynamics method is used as the main approach, we introduce multiobjective programming in the system dynamics model to develop an integrated model, which we call a system dynamics multiobjective programming model (SD-MOP), for the solution. The SD-MOP model not only provide better understanding of complex problems but also have considered the multiple objectives and also involve expert opinions in the decision. A general framework of the modeling process is shown in Figure 2. In a garden city transportation system decision process, a thorough analysis of the decision problem is conducted. Then, using the system dynamics (SD) approach, a causal loop diagram and detailed flow diagram are established. We run a series of MOPs to get the optimum value of those sensitive variables, and place these values into the SD model for simulation. Based on results of the SD-MOP integrated approach, different policy experiments are compared to choose the best route. If we are not satisfied with the result of the simulation, we can adjust the MOP models to yield better results; otherwise, the decision process is ended.

2.3. Basic Assumption

The basic assumptions of a garden city transportation system are as follows.

- (1) The main environmental pollution emissions we consider are CO_2 , excluding the exhausted gas from motorcycles.

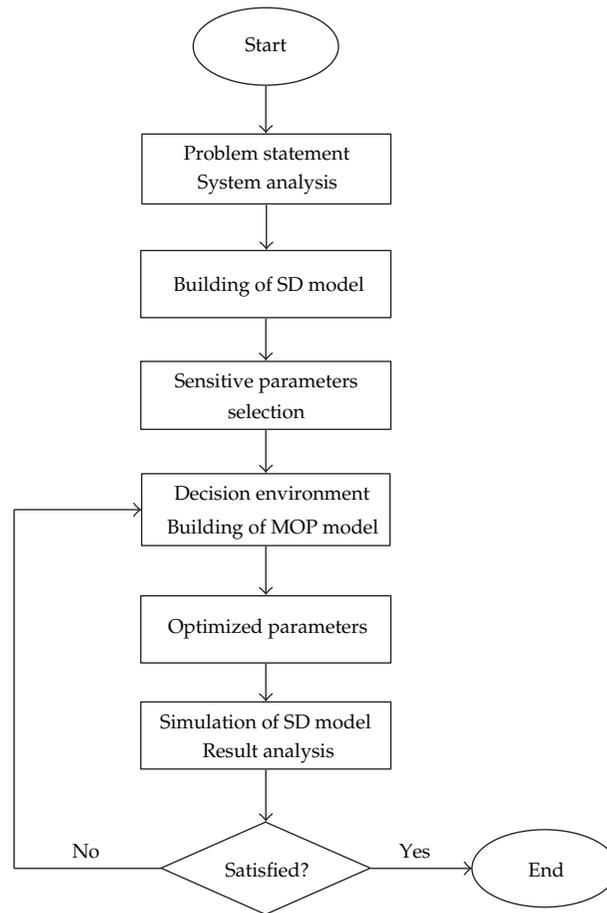


Figure 2: The general framework of the modeling process.

- (2) We consider private cars, buses, taxis, and rail as the four main transportation means that directly influence transportation congestion and ignore others such as bicycles, pedestrians, and others.
- (3) The influence of employment is ignored in the whole system so the employment is ignored.
- (4) We use gross domestic product (GDP) to measure the economic development.
- (5) The purpose of this study is to promote coordinated economic development, environmental protection, policy management, and population through the optimization of transportation construction and structure.

2.4. Index System

The transportation system is one of the most complex systems in modern city. Similarly, the analysis of transportation has been a vital element in world modern garden city. Factors analysis is an effective way to understand the structure and function of a system

well. According to the subsystems of garden city transportation system (Figure 1) and the characteristic of world modern garden city, we analyze the subdivision of each subsystem by selecting variables and influential factors synthetically based on the relative theoretical basis and the existing and our own research foundation, in principle of integrality, objectivity, scientificity, nonlinearity, practicality, and availability. Here, we list the main variables and influential factors of this model in Table 1 (variables and symbols in garden city transportation system). In order to facilitate our research and establish a mathematical model, we sort them according to the symbol of the name.

3. Modelling

Referring to the system description for garden city transportation structures, we construct a corresponding model. Firstly, the system dynamics general model is constructed. Secondly, a model is established, and the system dynamics model based on multiobjective optimization (SD-MOP) is developed. Finally, the model simulation method is analyzed.

3.1. System Dynamics Model

This Section is divided into two parts for a particular description of the modeling; firstly, the cause-effect relationship diagram, and, secondly, the stock and flow diagram, both of which are the two main steps when using system dynamics.

3.1.1. Cause-Effect Relationship Diagram

The SD model for the present study is developed for a transportation system. There are many variables in the subsystems occupying important positions in the system; thus, we build the cause-effect relationship diagram (see Figure 3) by incorporating the various features associated with the system. In this Figure, the arrows denote the cause-and-effect relationships and the plus and minus signs denote the positive and negative effects, respectively. The main feedback loops are given below:

- (1) economic development $\overset{+}{\rightarrow}$ total number of vehicles $\overset{+}{\rightarrow}$ transportation congestion $\overset{+}{\rightarrow}$ environmental pollution $\overset{-}{\rightarrow}$ economic development;
- (2) population $\overset{+}{\rightarrow}$ trip demand $\overset{+}{\rightarrow}$ total trips $\overset{+}{\rightarrow}$ transportation congestion $\overset{-}{\rightarrow}$ economic development $\overset{-}{\rightarrow}$ urban population;
- (3) economic development $\overset{+}{\rightarrow}$ infrastructure investment $\overset{+}{\rightarrow}$ road investment $\overset{+}{\rightarrow}$ road capacity $\overset{-}{\rightarrow}$ transportation congestion $\overset{-}{\rightarrow}$ economic development;
- (4) policy management $\overset{+}{\rightarrow}$ economic development $\overset{+}{\rightarrow}$ environmental pollution $\overset{-}{\rightarrow}$ policy intervention.

3.1.2. Stock and Flow Diagram

The causal relationship diagram emphasizes the feedback structure of the system, which, however, can never be comprehensive. We need to convert the causal relationship diagram

Table 1: Variables and symbols in garden city transportation system.

Sort	The meaning of variables	Variable units	Symbol
1	Road passenger capacity	Yuan	LEV _{CR}
2	Railway passenger capacity	Yuan	LEV _{CRW}
3	Exhaust emission	g	LEV _{EE}
4	Emission intensity	g	LEV _{EI}
5	Environmental pollution	L	LEV _{EP}
6	Fuel consumption volume	L/100 km	LEV _{FC}
7	Gross domestic product	Ten thousand	LEV _{GDP}
8	Investment of bus	Yuan	LEV _{I₁}
9	Investment of taxi	Yuan	LEV _{I₂}
10	Investment of railway	Yuan	LEV _{I₃}
11	Investment of infrastructure	Yuan	LEV _{II}
12	Investment of public transportation	Km	LEV _{IPT}
13	Investment of road	Yuan	LEV _{IR}
14	Length of road	Km	LEV _{LR}
15	Length of railway	Km	LEV _{LRW}
16	Number of buses	Ten thousand	LEV _{N₁}
17	Number of taxies	Ten thousand	LEV _{N₂}
18	Number of railbus	Ten thousand	LEV _{N₃}
19	Number of private cars	Ten thousand	LEV _{N₄}
20	Private car increment	Ten thousand	LEV _{PC}
21	Turnover of bus	Person/km	LEV _{T₁}
22	Turnover of taxi	Person/km	LEV _{T₂}
23	Turnover of railway	Person/km	LEV _{T₃}
24	Turnover of private car	Person/km	LEV _{T₄}
25	Traffic intensity	Person	LEV _{TI}
26	Traffic load	Person/yuan	LEV _{TL}
27	Total population	Ten thousand	LEV _{TP}
28	Travel volume of bus	Person/day	LEV _{V₁}
29	Travel volume of taxi	Person/day	LEV _{V₂}
30	Travel volume of railway	Person/day	LEV _{V₃}
31	Travel volume of private car	Person/day	LEV _{V₄}
32	Travel volume of public transport	Person/day	LEV _{VPT}
33	Travel volume of total trip	Person/day	LEV _{VT}
34	Rate of average trip	%	RAT _A
35	Transformation coefficient of bus	km/yuan	RAT _{C₁}
36	Transformation coefficient of taxi	km/yuan	RAT _{C₂}
37	Transformation coefficient of railway	km/yuan	RAT _{C₃}
38	Transformation coefficient of private car	km/yuan	RAT _{C₄}
39	Coefficient of discharge	g/L	RAT _D
40	Coefficient of economic decrease	Yuan/g	RAT _{ED}
41	Coefficient of emission factor	g/person	RAT _{EF}
42	Coefficient of environmental influence	No dimension	RAT _{EI}
43	GDP growth rate	%	RAT _{GDP}
44	Investment proportion of infrastructure	%	RAT _{II}
45	Investment proportion of public transportation	%	RAT _{IPT}
46	Investment proportion of road	%	RAT _{IR}

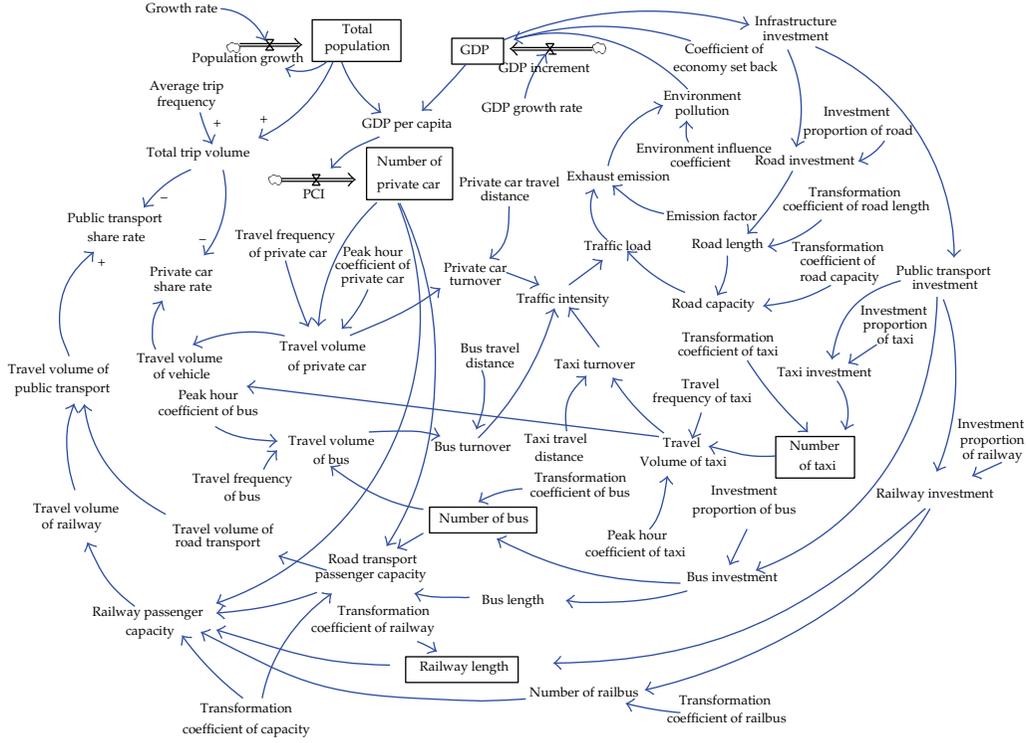


Figure 4: The stock and flow diagram of garden city transportation system.

into the stock and flow diagram that emphasizes the physical structure of the model, which tends to be more detailed than the causal loop diagram [28], to force us to think more specifically about the system structure. Figure 4 gives a detailed description, with the main formula as follows.

Through system dynamics modeling, we can get the first-order differential equations. The change rate of the turnover of bus, the $d(\text{LEV}_{T_1}(t))/dt$, is dependent on the stock of the turnover of bus $\text{LEV}_{T_1}(t)$, and there exists the basic stock O_{T_1} , which is subject to factors of transport planning, demand volume, and so on; besides, it would be effected by functioning time OT , which regularly means one year:

$$\frac{d(\text{LEV}_{T_1}(t))}{dt} = \frac{\text{LEV}_{T_1}(t) - O_{T_1}}{OT}. \quad (3.1)$$

Similarly, the differential equations of the turnover of the taxi, railway, and private car are

$$\begin{aligned} \frac{d(\text{LEV}_{T_2}(t))}{dt} &= \frac{\text{LEV}_{T_2}(t) - O_{T_2}}{OT}, \\ \frac{d(\text{LEV}_{T_3}(t))}{dt} &= \frac{\text{LEV}_{T_3}(t) - O_{T_3}}{OT}, \\ \frac{d(\text{LEV}_{T_4}(t))}{dt} &= \frac{\text{LEV}_{T_4}(t) - O_{T_4}}{OT}. \end{aligned} \quad (3.2)$$

Through the previous analysis, we get the main part of the links in the garden city transport system:

$$\begin{aligned}
\text{LEV}_{\text{GDP}}(t) &= \text{LEV}_{\text{GDP}}(t-1) \times \text{RAT}_{\text{GDP}}(t), \\
\text{LEV}_{\text{II}}(t) &= \text{LEV}_{\text{GDP}}(t) \times \text{RAT}_{\text{II}}(t), \\
\text{LEV}_{\text{IPT}}(t) &= \text{LEV}_{\text{II}}(t) \times \text{RAT}_{\text{IPT}}(t), \\
\text{LEV}_{\text{I}_1}(t) &= x_1 \cdot \text{LEV}_{\text{IPT}}(t), \\
\text{LEV}_{\text{N}_1}(t) &= \text{LEV}_{\text{I}_1}(t) \times \text{RAT}_{\text{C}_1}(t), \\
\text{LEV}_{\text{V}_1}(t) &= \text{PHF}_1 \times \text{TF}_1 \times \text{LEV}_{\text{N}_1}(t), \\
\text{LEV}_{\text{T}_1}(t) &= \text{TD}_1 \times \text{LEV}_{\text{V}_1}(t).
\end{aligned} \tag{3.3}$$

Till now, we obtained the turnover of bus LEV_{T_1} , similarly, the turnover of taxi, railway, and private car can be also described as LEV_{T_2} , LEV_{T_3} , and LEV_{T_4} , and the traffic intensity can be formulated as the following:

$$\text{LEV}_{\text{TI}}(t) = \text{LEV}_{\text{T}_1}(t) + \text{LEV}_{\text{T}_2}(t) + \text{LEV}_{\text{T}_3}(t) + \text{LEV}_{\text{T}_4}(t). \tag{3.4}$$

Further,

$$\begin{aligned}
\text{LEV}_{\text{TL}}(t) &= \frac{\text{LEV}_{\text{TI}}(t)}{\text{LEV}_{\text{CR}}(t)}, \\
\text{LEV}_{\text{EE}}(t) &= \text{LEV}_{\text{TL}}(t) \times \text{RAT}_{\text{EF}}, \\
\text{LEV}_{\text{EP}}(t) &= \text{LEV}_{\text{EE}}(t) \times \text{RAT}_{\text{EI}}, \\
-\text{LEV}_{\text{GDP}}(t) &= -\text{LEV}_{\text{EP}}(t) \times \text{RAT}_{\text{ED}},
\end{aligned} \tag{3.5}$$

where it can indicated that the irrational structure of transportation can increase the environmental pollution and ultimately decrease the development of economy to a certain extent. Meanwhile,

$$\begin{aligned}
\text{LEV}_{\text{TP}}(t) &= \text{LEV}_{\text{TP}}(t-1) + \text{LEV}_{\text{TP}}(t-1) \times \text{RAT}_{\text{TP}}(t), \\
\text{LEV}_{\text{VT}}(t) &= \text{LEV}_{\text{TP}}(t) \times \text{RAT}_{\text{TF}}, \\
\text{RAT}_{\text{SR}_1} &= \frac{\text{LEV}_{\text{VPT}}(t)}{\text{LEV}_{\text{VT}}(t)}, \\
\text{RAT}_{\text{SR}_2} &= \frac{\text{LEV}_{\text{V}_4}(t)}{\text{LEV}_{\text{VT}}(t)}.
\end{aligned} \tag{3.6}$$

Through this circulate series of formulation, each variable (the standard variable and rate variable) is defined, thus, building the foundation of our model.

3.2. Multiobjective Programming Model

The purpose of multiobjective programming (MOP) is to maximize (or minimize) different multiobjective functions under a set of constraints, which is suitable for decision making in systems which have two or more goals [29]. According to the analysis of the system above, the optimization of a garden city transportation structure needs to consider the economic, social, and environmental subsystems and the transportation structure together to maximize the final benefits. Therefore, a multiobjective method can be used to solve this problem. In this Section, we will apply a multiobjective optimization model into the stock and flow diagram to measure the most optimal transportation structure to invest.

3.2.1. Objective Function

As usual, the objective is to pursue the maximal economic and, social benefit with minimal environmental pollution. Here we list our three main objective functions.

(1) Maximal Gross Domestic Product (GDP)

Economy is an important part involved in garden city transportation system, and, often, we use GDP to measure the level of it, the higher the GDP we produce, the better we operate our country and more investment on transportation be conducted and the system develops better:

$$\max f_1 = \text{LEV}_{\text{GDP}} - M_1 \sum_{i=1}^4 \text{LEV}_{\text{EP}_i} \cdot \text{RAT}_{\text{ED}_i}, \quad (3.7)$$

here, the LEV_{GDP} represents the quantity of GDP, and $\text{LEV}_{\text{EP}_i} \cdot \text{RAT}_{\text{ED}_i}$ is the economy decrease caused by environment pollution of each transportation means. Since LEV_{EP_i} represents the environment pollution, while GDP is dimensional, we add an M_1 to the balance to make them under the same unit.

(2) Less Environment Pollution

As the economy develops, the public consciousness of environmental protection is aroused. The automobile exhaust emission occupies most parts in air pollution, so the optimal the transportation structure combination, the minimal air pollution and environmental damage. To achieve this, we must guarantee the least exhaust emissions:

$$\min f_2 = \sum_{i=1}^4 \text{LEV}_{\text{EE}_i} - \delta_1 f_1, \quad (3.8)$$

where LEV_{EE_i} represents the level of exhaust emission, and $\delta_1 f_1$ represents the environment pollution reduction resulting from economy growth investment.

(3) *More Social Benefits*

Social benefit is also an important aspect. Because transportation system can make people's life more convenient; if the system is not operating well, there will not be sufficient supply for people to travel. Therefore, it needs more turnover of each transport means to bear people's travel demand, which is a criterion to measure the transport capacity:

$$\max f_3 = \sum_{i=1}^4 \text{LEV}_{T_i} - M_2 \delta_2 f_1 - M_3 \delta_3 f_2. \quad (3.9)$$

LEV_{T_i} is the turnover of each means of transportation and $\delta_2 f_1$ and $\delta_3 f_2$ represents the negative influence on turnover of transport means from the economy growth investment and environment pollution respectively. M_2 and M_3 are an equivalent used to balance different units.

3.2.2. *Constraints*(1) *Total Transportation Structure Proportion Constraint*

We assumed that there are only four means of transportation in the system, thus making the sum total 1:

$$\sum_{i=1}^4 x_i = 1. \quad (3.10)$$

(2) *Investment Constraint*

Because plans have been made in the government 5-year plans the transportation structures and therefore expenditures have already been determined. Thus, for each means of transport considered here there is a maximum and minimum ranges:

$$\begin{aligned} a_{11} &\leq x_1 \leq a_{12}, \\ a_{21} &\leq x_2 \leq a_{22}, \\ a_{31} &\leq x_3 \leq a_{32}, \\ a_{41} &\leq x_4 \leq a_{42}, \\ b_1 &\leq x_1 + x_2 + x_3 \leq b_2, \end{aligned} \quad (3.11)$$

x_1, x_2, x_3, x_4 represent each means of transportation, here, $a_{11}, a_{21}, a_{31}, a_{41}$ represent the lower limit of proportion, while the $a_{12}, a_{22}, a_{32}, a_{42}$, the upper limit. b_1 is the lowest proportion of public transport, while b_2 is the upper limit.

(3) *Ratio Constraint*

There are two kinds of share rate in this system, the public transportation share rate and private car share rate, both of them are between 0 ~ 1, and the sum of them is equal to 1:

$$\begin{aligned} 0 < \frac{LEV_{VPT}}{LEV_{VT}} < 1, \\ 0 < \frac{LEV_{V_4}}{LEV_{VT}} < 1. \end{aligned} \quad (3.12)$$

(4) *Intensity Constraint*

Usually, the emission intensity index decreases along with technological progress and economic growth. The emission intensity of this year is expected to be smaller than that of the last year. Therefore, the emission intensity has an upper limit and decreases every year:

$$\frac{\sum_{i=1}^4 LEV_{EE_i}}{f_1} \leq \sum_{i=1}^4 LEV_{EI_i} (1 - c_1), \quad (3.13)$$

where $LEV_{EE_i}(t - 1)$ represent the exhaust emission and $LEV_{EI_i}(t - 1)$ represents emission intensity of last year, and c_1 is the the average rate of decrease required.

Similarly, the intensity of road occupation to economy also decreases:

$$\frac{f_3}{f_1} \leq \sum_{i=1}^4 LEV_{TI_i} (1 - c_2), \quad (3.14)$$

where $LEV_{TI_i}(t - 1)$ represents road occupation intensity and c_2 is the the average rate of decrease required.

From this, we get (3.16) as follows:

$$\begin{aligned} \max f_1 &= LEV_{GDP} - M_1 \sum_{i=1}^4 LEV_{EP_i} \cdot RAT_{ED_i} \\ \min f_2 &= \sum_{i=1}^4 LEV_{EE_i} - \delta_1 f_1 \\ \max f_3 &= \sum_{i=1}^4 LEV_{TI_i} - M_2 \delta_2 f_1 - M_3 \delta_3 f_2 \end{aligned} \quad (3.15)$$

$$\text{s.t.} \left\{ \begin{array}{l}
\sum_{i=1}^4 x_i = 1 \\
a_{11} \leq x_1 \leq a_{12} \\
a_{21} \leq x_2 \leq a_{22} \\
a_{31} \leq x_3 \leq a_{32} \\
a_{41} \leq x_4 \leq a_{42} \\
b_1 \leq x_1 + x_2 + x_3 \leq b_2 \\
0 < \frac{\text{LEV}_{\text{VPT}}}{\text{LEV}_{\text{VT}}} < 1 \\
0 < \frac{\text{LEV}_{V_4}}{\text{LEV}_{\text{VT}}} < 1 \\
\frac{\sum_{i=1}^4 \text{LEV}_{\text{EE}_i}}{f_1} \leq \sum_{i=1}^4 \text{LEV}_{\text{EI}_i} (1 - c_1) \\
\frac{f_3}{f_1} \leq \sum_{i=1}^4 \text{LEV}_{\text{TI}_i} (1 - c_2) \\
e_1 \leq \delta_i \leq e_2 \\
0 \leq a, b, c, e \leq 1, \quad i = 1, \dots, 4.
\end{array} \right. \quad (3.16)$$

3.3. Solution Method

In this Section, we make use of the ideal point method proposed by Yingming et al. [30]; Rakowska et al. [31]; and William [32] to resolve the multiobjective problem (3.16) with crisp parameters [33]. If the policy maker can firstly propose an estimated value \bar{F}_i for each objective function $f_i(x)$ such that

$$\bar{F}_i \geq \max_{x \in X} f_i(x), \quad i = 1, 2, 3, \quad (3.17)$$

where $X = \{x \mid x \in X\}$, X is the collection range of constraints, and then $\bar{\mathbf{F}} = (\bar{F}_1, \bar{F}_2, \bar{F}_3)^T$ is called the ideal point, especially if $\bar{F}_i \geq \max_{x \in X} f_i(x)$ for all i , we call $\bar{\mathbf{F}}$ the most ideal point.

The basic theory of the ideal point method is to take an especial norm in the objective space \mathbf{R}^m and obtain feasible solution x so that the objective value approaches the ideal point $\bar{\mathbf{F}} = (\bar{F}_1, \bar{F}_2, \bar{F}_3)^T$ under the norm distance, that is, to seek the feasible solution x satisfying

$$\min_{x \in X'} u(f_i(x)) = \min_{x \in X'} \|f_i(x) - \bar{\mathbf{F}}\|. \quad (3.18)$$

Next, we take the p -mode function to describe the procedure for solving the problem (3.16).

Step 1. Find the ideal point. If the decision maker can give an ideal objective value satisfying condition (3.17), the value will be considered the ideal point. However, decision makers do

not know how to estimate the objective value, so we can get the ideal point by solving the following programming problem:

$$\begin{aligned} \max \quad & f_i(x) \\ \text{s.t.} \quad & x \in X. \end{aligned} \quad (3.19)$$

Then the ideal point $\bar{F} = (\bar{F}_1, \bar{F}_2, \bar{F}_3)^T$ can be fixed by $\bar{F}_i = f_i(x^*)$, where x^* is the optimal solution of problem (3.19).

Step 2. Fix the weight. The method of selecting the weight is referred to in much research that interested readers can consult these. We usually use the following function to fix the weight:

$$\omega_i = \frac{\bar{F}_i}{\sum_{i=1}^3 \bar{F}_i}. \quad (3.20)$$

Step 3. Construct the minimal distance problem. Solve the following single-objective programming problem to obtain an efficient solution to problem (3.16):

$$\begin{aligned} \min \quad & \left[\sum_{i=1}^m \omega_i |f_i(x) - \bar{F}_i|^t \right]^{1/t} \\ \text{s.t.} \quad & x \in X, \end{aligned} \quad (3.21)$$

usually, we take $t = 2$ to compute it.

4. A Case Study

In this section, we choose Chengdu, first city that advocates to “being a world modern garden city,” in China as our application to verify the approach in the previous section. we apply the data and parameter values of Chengdu into the system dynamics model. A system simulation was performed using the simulation software VENSIM and the data from 2010 as initial conditions, time = 0. Our simulation spans 11 years, from 1 to 11, and results in data analysis for the years 2010 to 2020 and we depict the main pattern in figures.

4.1. Regional Situation

As a general transportation hub for western China, Chengdu is an important nexus city linking up China to mid-Asia, south Asia, west Asia, and Europe. Located in the middle of Sichuan province in southwest China, Chengdu covers a total land area of 12121 square kilometers, with its central downtown area extending for approximately 350 square kilometers. With its name and location kept unchanged for more than 2300 years, the city’s history traces back far and the culture reaches wide. As the main hub for western transportation and the most developed city in the southwest China, with the nature advantages, the proposition of the objective of a “world modern garden city” is necessary

and surely no accident. Chengdu is blessed with unsurpassed resources, and the nature, humanity, and history of Chengdu make it well qualified for garden city construction. In late 2009, the city committee and government made the development of a “world modern garden city” its historic positioning and long-term target based on in-depth research, sufficient analysis, and extensive public participation in the notion. It presents an attempt to capitalize on the historic opportunities generated by the prosperity of China to further the urban-rural integration and push along the strategic transformation of growth models so that the city can better contribute to the new round of opening-up and development activities in western China and to the province’s strategic move to become the top driving force in the development of western China. However, in the way of garden city construction, the transport problems have become increasingly severe. We have to deal with the traffic problems as an ordinary city and as a particular problem emerged in the construction of world modern garden city, which is brand new for us.

4.2. Simulation Results and Analysis

We collect the parameter statistics by studying the garden city transportation system and analyzing the flows of processing technique and show the results in Table 2. The settled values for the substance transforming rates, and some settled parameters in the system dynamic model, are mainly based on the administration annual report for the region: and National Statistical Bureau [34, 35], Ministry of Transport and Communication [36], Chengdu Bureau of Statistics and the planning reports: Chengdu Twelfth Five Years of planning [37], and National Twelfth Five Years of planning [38] on correlative industries and the present market situation. The settled values were obtained via equilibration, linearity regression, index smoothness, and other related mathematical models based on the principles like relativity, comparability, scientificity, and comprehensiveness. We define the parameters used to describe and analyze the system, and the parameters of the transportation system are presented in Table 2.

In order to achieve the government’s goal, a multiobjective optimization problem incorporating the decision makers’ preferences is formulated. The multiobjective model is based on Model (3.16). There are some parameters which are determined by the decision maker of local government. In the current case, parameters such as a , b , c , d have to be given according to the preferences of decision maker. The decision maker can provide different values and decide which solutions are adopted by comparing the solutions. The decision maker is encouraged to give probable numbers to express their preference. With this method, we obtained three different solutions as shown in Table 3 (control variables) for different weights considered for the objective functions, among which the current program presents the current situation of transport proportion. In this table, the different proportions of transport means show that from the current program to optimization program 3, the proportion of public transport is increasing, especially the bus and railway, while the private car decreases dramatically, and the number of taxi decreases slightly. Finally, these numbers and cases will be used as control variables in transportation system dynamics modeling to operate along with the initial data. We use the four groups of figures (Table 3) to predict the coming 10 years, and, in turn, suggest actions to improve the present situation.

The results after the system dynamics modeling are shown from Figure 5 to Figure 10. As the system is simulated, six variables are selected for observation which are classified in three groups: the transport structure represented by the public transport share rate and railway passenger capacity (Figures 5 and 6), the road use situation represented by

Table 2: The value of parameters.

Symbol	Value	Units
RAT_A	82	%
RAT_{C_1}	96.94	%
RAT_{C_2}	94.40	%
RAT_{C_3}	92.30	%
RAT_{C_4}	90.50	%
RAT_D	3.9175	g/L
RAT_{ED}	-0.29	Yuan/g
RAT_{EF}	1.5	g/person
RAT_{EI}	0.6	No dimension
RAT_{GDP}	14.70	%
RAT_{II}	12	%
RAT_{IPT}	40	%
RAT_{IR}	60	%
PHF_1	0.3242	Person/times
PHF_2	0.2260	Person/times
PHF_3	0.3065	Person/times
PHF_4	0.4006	Person/times
TD_1	10.545	km/time
TD_2	9.413	km/time
TD_3	12.413	km/time
TD_4	10.517	km/time
TF_1	11.37	Times
TF_2	10.5	Times
TF_3	12.28	Times
TF_4	4.16	Times

Table 3: Control variables.

	Current program	Optimization program 1	Optimization program 2	Optimization program 3
x_1	0.3830	0.3000	0.3600	0.1750
x_2	0.1036	0.2000	0.1375	0.2096
x_3	0.4080	0.2750	0.3400	0.2658
x_4	0.1054	0.2250	0.1625	0.3496

the traffic intensity and public transport volume (Figures 7 and 8), and the environmental circumstances represented by the exhaust emissions and environment pollution (Figures 9 and 10). From Figure 5, the public transport share rate is increasing, if no changes had taken place, the public transport share rate will remain the same as the current situation. And we can see that optimization 3 is the highest in the coming 10 years, and optimization 2 is higher than optimization 1 in the first several years, and all the three optimizations show significant improvement than continuing with the current situation. Figure 6 describes that the railway passenger capacity will be greatly strengthened if optimization cases are adopted. In terms of the road use situation, Figure 7, traffic intensity means the higher the intensity, the more utilization of the road, and Figure 8 presents travel volume of public transport is always increasing, and indicates the demand of transport. Both of them show that optimization 3 is the best choice. Lastly, in terms of the environmental circumstance, both Figures 9 and 10

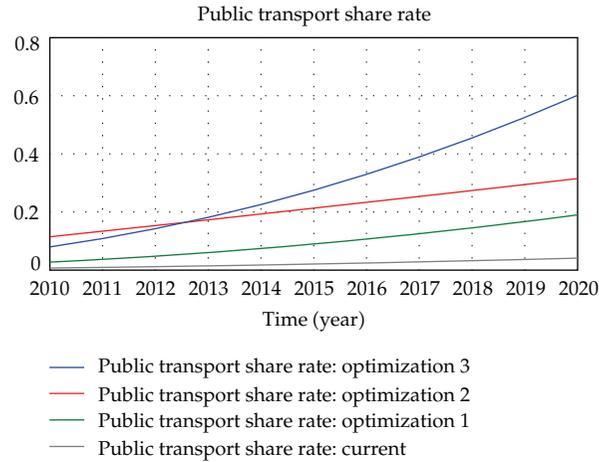


Figure 5: Public transport share rate.

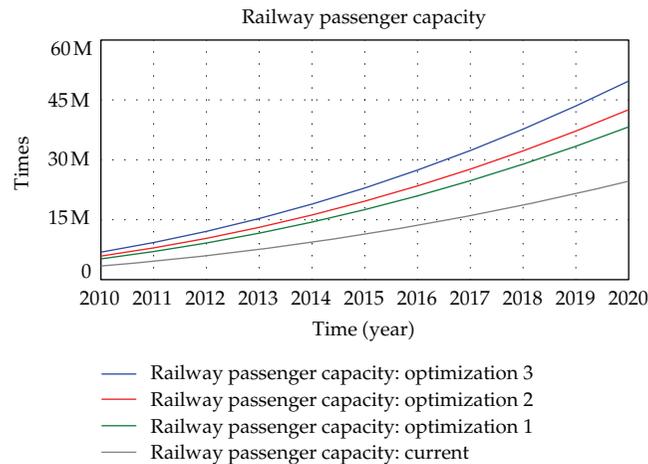


Figure 6: Railway passenger capacity.

are critical factor in assessing environment contamination. Most of the exhaust emission contributes to the environment pollution, and there is a linear relationship between them. Generally speaking, from analysis of these 6 graphs we observe that the most suitable case is optimization 3, which in all cases offers better progress towards the goal of a garden city than the current situation.

5. Proposals

From an analysis of the results and in consideration of local conditions, suggestions are made to find a feasible solution to the transportation development in garden city. To achieve a continual optimization of the transportation structure, low-carbon transportation development needs to be promoted through the formulation of relevant policy by the local government.

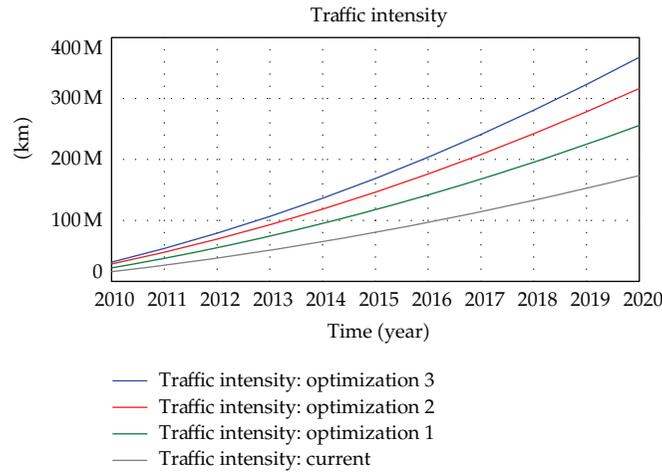


Figure 7: Traffic intensity.

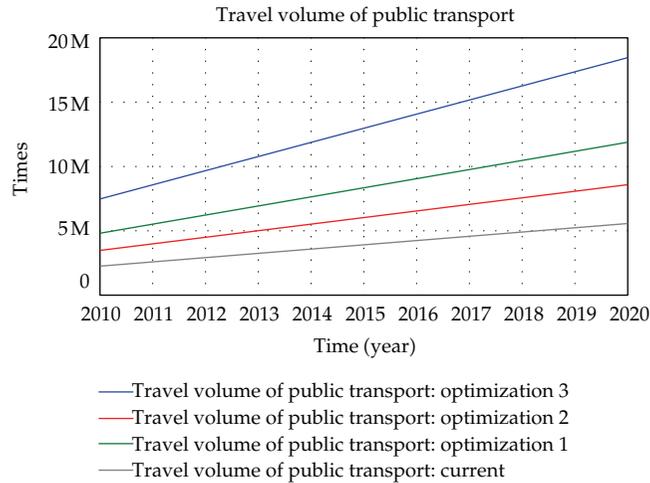


Figure 8: Travel volume of public transport.

(1) Transportation Structure Optimization

The adjustment of the transportation structure is a long complicated process which cannot be achieved through a single policy. A previous simulation confirms that if the government ensures the preferential development of the public transport, the situation can get better. Firstly, develop the loop line. The loop line and radiation transport route can associate the central city with the countryside areas, which takes a great advantage of the plain landscape, as well as, the standard model of garden city transportation. Secondly, the subway construction. It is convinced that the railway transport is the context to open garden city construction. With the 1st line of Chengdu Metro operating well and its notable benefit, new subways should be constructed to spread further and ultimately to the whole city. Thirdly, construct the bus rapid transit (BRT), which is an important part in mass rapid public transit. Here, the mass rapid public transit is an resources saving and socially beneficial which

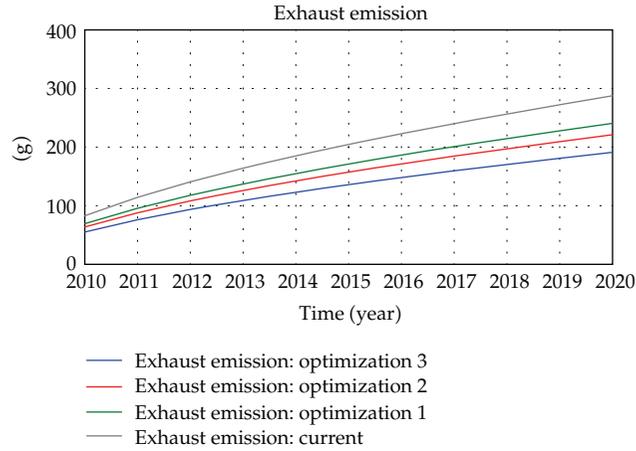


Figure 9: Exhaust emission.

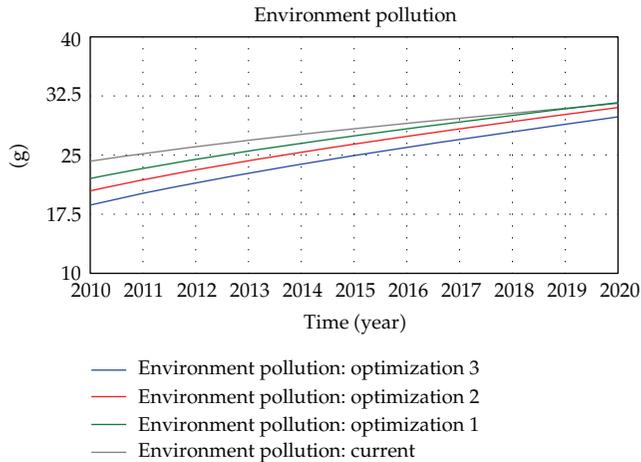


Figure 10: Environment pollution.

includes BRT, subway, light railway, and others. The bus rapid transit (BRT) integrates the bus technology, intelligent traffic system with rail transit operations management mode into a relatively low-cost mass transit mode, which is regarded as a revolutionized solution of public transport by International Energy Agency. Fourthly, restrict development of private cars. As the private cars are at an absolute disadvantage in transport area possession, energy consuming, and exhaust emissions, measures should be taken to restrict it such as purchase limiting, levying a tax based on the emissions to the environment, and reducing the supply of oil.

(2) Promoting the Land Use Mode

There is a close relationship between land use planning and transport construction. Land use planning is crucial to the whole transportation system. As our simulation model shows, land use influences the transport infrastructure such as road length, road capacity, and

the investment proportion of road. According to the actual land situation of Chengdu, firstly, emphasis on the mixed use of land and the various functional complement with each other is applied in order to improve the overall efficiency of central district. Secondly, newly developed areas should conduct preliminary transport planning before the land use plan. Thirdly, the public transport guide land use mode needs to be developed to shorten the travel distance between bus stops and increase the operating speed. Fourthly, controlling the diversity of land use intensity along the road lines to maintain traveling speed and raise land use intensity and efficiency. Fifthly, attention should be paid to the ecological construction of land. The government should transfer the distant and relatively large land parcels to farmland to develop urban agriculture which can improve the environmental quality, and suppress the unlimited extension of city construction land use scale, for example, the Shahe, Sansheng, and Shiling park. Lastly, free up the original lane space to provide for the new bus station and parking place, at the same time, set up new bicycle squares in the city center.

(3) Strengthening the Low-Carbon Consciousness

Because there is a high value put on environmental protection in our model, it is necessary to examine and weigh the optimization using environmental indexes. A green travel consciousness needs to be developed which would focus on the sustainable development of the urban inhabitable environment. Thus, both the walking and alternative means need to be promoted along with the low-carbon, safe, comfortable, and low-pollution public transport. It is an effective way to introduce the low-carbon consciousness into primary school classroom education, which cannot only guide the students to establish the right consumption concept, but also to foster their socially responsible manner and has a profound effect on low-carbon transportation construction. This promotion of green transportation will lower the dependence on motor vehicles and encourage people to use hybrid or clean fueled vehicles which would also satisfy one of the measures proposed by the State Council for reducing greenhouse gas emissions and conserving energy.

6. Conclusion

Great many cities are experiencing traffic problems in the process of city development, including the garden city. Many researchers did not conclude all the factors in the transport system when estimating it. Therefore, to address the problems, a methodology for the analysis of the whole transportation system by adjustment of transport structure was outlined. Particularly, we develop a system dynamics and multiobjective programming integrated model (SD-MOP) to simulate different results with different proportions of transport means. The system dynamics model describes the relationships between the economy, environment, traffic mode traffic congestion, and the policy management. A multiobjective programming model helps policy makers to make choices according to their preferences. In the case study, a representative city, Chengdu, a world modern garden city was chosen. Various scenarios and different optimization cases were simulated to show the future trends of the transportation system. According to the simulation results, we propose the reasonable pieces of advice on transportation structure and transportation development mode in world modern garden city.

It is of great significance to study transportation system by system dynamics integrated with multiobjective optimization model. There is much scope to expand this field of research into the future. There may exist some omissions in this system, future research will

focus on establishing a more complex transportation system which considers more factors and deals with other optimizations.

Acknowledgments

This research is supported by the Key Program of NSFC (Grant no. 70833005), Chinese Universities Scientific Fund (Grant no. 2010SCU22009), and the Key Project of Philosophy and Social Sciences Research, Ministry of Education of China (Grant no. 08JHQ0002). The authors would like to thank the anonymous referees for their insightful comments and suggestions to improve this paper, as well as Uncertainty Decision-Making Laboratory and Low-Carbon Technology and Economy Research Center of Sichuan University for helpful comments and discussion.

References

- [1] C. Cooley, "The theory of transportation," *Economics Association*, vol. 9, no. 3, pp. 1–148, 1894.
- [2] F. Moavenzadeh and D. Geltner, *Transportation, Energy and Economic Development: A Dilemma in the Developing World*, Elsevier, New York, NY, USA, 1984.
- [3] W. Owen, *Transportation and World Development*, Johns Hopkins University Press, Baltimore, Md, USA, 1987.
- [4] International Energy Agency (IEA), *Key World Energy Statistics*, International Energy Agency (IEA), 2010b.
- [5] T. Petri, B. David, L. Jyrki, V. Jarmo, and W. Risto, "Energy and transport in comparison: Immaterialisation, dematerialisation and decarbonisation in the EU15 between 1970 and 2000," *Energy Policy*, vol. 35, no. 1, pp. 433–451, 2007.
- [6] S. Dominic, "Transport energy efficiency in Europe: temporal and geographical trends and prospects," *Journal of Transport Geography*, vol. 15, no. 5, pp. 343–353, 2007.
- [7] S. Frank, A Technical Review of Urban Land use. Transportation Models as Tools for Evaluation Vehicle Travel Reduction Strategies. The Office of Environmental Analysis and Sustainable Development U.S Department of Energy, 1994, <http://www.thepep.org/ClearingHouse/docfiles/A%20Technical%20Review%20of%20Urban%20Land%20Use-Transportation%20Models%20as%20Tools%20for%20Eva.pdf>.
- [8] T. Messenger and R. Ewing, "Transit-oriented development in the sun belt," *Transportation Research Record*, no. 1552, pp. 145–153, 1996.
- [9] L. Martin, "Means of transportation to work and overweight and obesity: a population-based study in southern Sweden," *Preventive Medicine*, vol. 46, no. 1, pp. 22–28, 2008.
- [10] S. Komei, "Transportation system change and urban structure in two-transport mode setting," *Journal of Urban Economics*, vol. 25, no. 3, pp. 346–367, 1989.
- [11] S. Jian, L. Qiong, and P. Zhongren, "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.
- [12] H. Wang, P. Zhou, and D. Q. Zhou, "An empirical study of direct rebound effect for passenger transport in urban China," *Energy Economics*, vol. 34, no. 2, pp. 452–460, 2012.
- [13] W. Xiao, "Discussions on some issues of the eleventh five-year plan for transport of China," *Journal of Transportation Systems Engineering and Information Technology*, vol. 6, no. 6, pp. 1–5, 2006.
- [14] C. Bofeng, Y. Weishan, C. Dong, L. Lancui, Z. Ying, and Z. Zhansheng, "Estimates of China's national and regional transport sector CO₂ emissions in 2007," *Energy Policy*, vol. 41, pp. 474–483, 2012.
- [15] E. Howard, G. H. Peter, D. Hardy, and W. Colin, *To-Morrow: A Peaceful Path to Real Reform*, Routledge, London, UK, 1898.
- [16] E. Howard, *Cities of to-Morrow*, MIT Press, Cambridge, Mass, USA, 1965.
- [17] P.D. Kieran and A. S. Laurie, "Managing congestion, pollution, and pavement conditions in a dynamic transportation network model," *Transportation Research Part D*, vol. 3, no. 2, pp. 59–80, 1998.
- [18] W. Jifeng, L. Huapu, and P. Hu, "System dynamics model of urban transportation system and its application," *Journal of Transportation Systems Engineering and Information Technology*, vol. 8, no. 3, pp. 83–89, 2008.

- [19] Ü. Füsün, Ö. Şule, Y. Ilker Topçu, A. Emel, and K. Özgür, "An integrated transportation decision support system for transportation policy decisions: the case of Turkey," *Transportation Research Part A*, vol. 41, no. 1, pp. 80–97, 2007.
- [20] S. Anand, P. Vrat, and R. P. Dahiya, "Application of a system dynamics approach for assessment and mitigation of CO₂ emissions from the cement industry," *Journal of Environmental Management*, vol. 79, no. 4, pp. 383–398, 2006.
- [21] K. S. Ali, B. Yaman, and Y. Orhan, "Environmental sustainability in an agricultural development project: a system dynamics approach," *Journal of Environmental Management*, vol. 64, no. 3, pp. 247–260, 2002.
- [22] F. Andrew, "System dynamics and the energy industry," *Encyclopedia of Energy*, vol. 40, no. 1, pp. 809–818, 2004.
- [23] L. C. Wadhwa and Y. M. Demoulin, "A system approach to regional economic and transport planning in Australia," in *Proceedings of the 2nd International Symposium on Large Engineering Systems*, pp. 95–100, Ontario, Canada, 1978.
- [24] Y. Tanaboriboon, *A transportation planning strategy for the Bangkok metropolitan area [Ph.D. thesis]*, Virginia Polytechnic Institute and State University, Virginia, Va, USA, 1979.
- [25] J. D. Sterman, "Learning in and about complex systems," *System Dynamics Review*, vol. 10, no. 2-3, pp. 291–330, 1994.
- [26] X. Jiuping, D. Rentao, and D. W. Desheng, "On simulation and optimization of one natural gas industry system under the rough environment," *Expert Systems with Applications*, vol. 37, no. 3, pp. 1854–1862, 2010.
- [27] Y. N. Yu, *Electric Power System Dynamics*, Academic Press, New York, NY, USA, 1983.
- [28] B. Tung and L. Claudia, "Supporting cognitive feedback using system dynamics: a demand model of the Global System of Mobile telecommunication," *Decision Support Systems*, vol. 17, no. 2, pp. 83–98, 1996.
- [29] D. Kalyanmoy and T. Santosh, "Omni-optimizer: a generic evolutionary algorithm for single and multi-objective optimization," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1062–1087, 2008.
- [30] W. Yingming, Y. Jianbo, X. Dongling, and C. Kwaisang, "The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees," *European Journal of Operational Research*, vol. 175, no. 1, pp. 35–66, 2006.
- [31] J. Rakowska, R. T. Haftka, and L. T. Watson, "Tracing the efficient curve for multi-objective control-structure optimization," *Computing Systems in Engineering*, vol. 2, no. 5-6, pp. 461–471, 1991.
- [32] K. S. William, "Multiobjective decision analysis with engineering and business applications," *Engineering Geology*, vol. 19, no. 4, pp. 289–291, 1983.
- [33] J. Xu and L. Yao, *Random-Like Multiple Objective Decision Making*, vol. 647 of *Lecture Notes in Economics and Mathematical Systems*, Springer, Heidelberg, Germany, 2011.
- [34] National Statistical Bureau, *Statistical Yearbook [M]*, Press of China, Beijing, China, 2010.
- [35] National Statistical Bureau, *China Energy Statistical Yearbook [M]*, Statistical Press of China, Beijing, China, 2010.
- [36] Ministry of Transport and Communication, *China Transport Statistical Yearbook [M]*, People's Communication Press, Beijing, China, 2010.
- [37] Chengdu, Economic and Social Development of the Twelfth Five Years of planning [M], 2011.
- [38] National Economic and Social Development of the Twelfth Five Years of planning [M], 2011.

Research Article

A Signal Coordination Control Based on Traversing Empty between Mid-Block Street Crossing and Intersection

**Changjiang Zheng,¹ Genghua Ma,² Jinhua Wu,¹
Xiaoli Zhang,¹ and Xiaolei Zhang¹**

¹ College of Civil and Transportation Engineering, Hohai University, Nanjing 210098, China

² College of Harbour, Coastal and Offshore Engineering, Hohai University, Nanjing 210098, China

Correspondence should be addressed to Jinhua Wu, wujinhua115@126.com

Received 29 June 2012; Revised 6 September 2012; Accepted 11 September 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Changjiang Zheng 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 solve the problem in pedestrian Mid-Block street crossing, the method of signal coordination control between mid-block street crossing and intersection is researched in this paper. The paper proposes to use “distance-flow rate-time” graph as the tool for building coordination control system model which is for different situations of traffic control. Through alternating the linear optimization model, the system outputs the distribution of signal timing and system operational factors (delays in vehicles and mid-block street crossing). Finally, taking one section on the Taiping North Road in Nanjing as an example, the signal coordination control is carried out. And the results which are delays in the vehicles and mid-block street crossing are compared to those in the current distribution of signal timing.

1. Introduction

There are two kinds of representation methods in studying discrete feature of vehicles flow on the road, normal distribution function-Pacey method (1956) and geometric distribution function-Robertson method (1969) [1]. A. F. Rumsey and M. G. Hartley carried out computer simulations for comparison tests about these two distribution function methods proposed. They found although arrival rates of vehicle flow in downstream crossing which are calculated by those two different types of distribution functions are different, it is only a little effect on signal timing design. At present, the empirical formula that D. I. Robertson established based on the geometric distribution assumption is applied widely.

Studying pedestrian traffic characteristics, the American scholar JTFruit (1973) has researched minutely on pedestrian walking speed, density, and traffic flow, pedestrian spatial

characteristics, and their relationships in this doctoral thesis, Pedestrian Planning and Design. He also proposed recommended value of the level of service of the sidewalk; Japanese scholars, Hosoi et al. have conducted a study on pedestrians behaviour in the crowd [2]. Wang et al. have carried out simulation and analysis of driver's deceleration and acceleration behaviour under different driving situations after identified the key safety-related parameters [3]. Wang et al. investigated driver's car-following safety behaviour in more detail and then developed an integrated safety approaching model and simulation procedures for devising traffic accident solution and evaluating urban traffic at the microscopic level [4]. Data of nearly 3,500 pedestrian crossing operations were collected at 27 crosswalks in the Greater Amman Area. Statistical analyses were conducted to reveal which factors significantly contribute to pedestrians' speeds [5].

Many foreign scholars have conducted many successful researches and applications on coordination control of traffic signal in urban major trunk roads. In early period, through theoretical analysis, scholars did research on coordination control of traffic signal using the time-distance diagram for timing-type liner control system [6].

With the application of computers, people could use a computer to get program of control which might be difficult to be achieved manually. "Off-line" and "on-line" are the two computer methods to be used to calculate coordinate liner control system. Off-line method is prepared for computing software which is in the light of some kind of optimized principles. Liner timing methods can be developed by computer. Then these control programs are set to the signal control machine which makes the signal lamps be in operation in accordance with the timing program set from time to time. With these methods, computing software becomes the point which scholars focus on. Among these softwares, MAXBAND and PASSER II are more representative [7]. In on-line method, the timing Programs of liner control system are obtained by computing software, and input data computing software needed is directly received from the vehicle detector through computer. Also the signal lights operate under computer. It adopts two kinds of methods which are "Type of Timing programs Selection" and "Type of Timing Programs Production." Coordination control system is represented by the SCOOT, SCAT, UTCS, and RT-TRANCS.

Sun et al. establish a model of motorist-pedestrian. The model uses a binary logit approach; multiattribute regression analyses are performed to capture the decision making process of the pedestrian [8]. Wolshon and Taylor analyze the differences in certain delay parameters which would occur as a result of implementing SCATS signal control [9]. Trabia et al. present the design and evaluation of a fuzzy logic traffic signal controller for an isolated intersection [10]. In adaptive method, the phase, and green signal time of single intersection were set by fuzzy controller. Then the average delay of vehicles was taken as objective function, the green time of every single intersection was dynamically adjusted, and the phase shift was optimized through genetic algorithm [11].

Domestic and foreign scholars have did some research and application on coordination control between pedestrian mid-block street crossing and intersections, but not conducting systemic and deep studies and not forming a complete research results. Foreign scholars studied them from the perspective of speed discreteness. Domestic researches were different from them. In the view of concentration, Canqi and Peikum (1996) proposed diffusion models of traffic platoon, diffusion models of traffic head, and diffusion models of traffic rear [12]. Domestic scholars have studied various aspects of pedestrian walking characteristics. They conduct some researches on pedestrians' walking speed, stride, and occupying space of walking, and so forth.

In actual application, domestic liner control system usually uses foreign technology and methods. Beijing has conducted liner control test which had good effect. Then it was gradually applied in some typical road sections. Nanjing's urban traffic control system was completely and successfully developed by self-localization. It was fit for the compound traffic conditions of our country, low-density road network, and special city road condition with great disparity gap between two junctions. Equipping three kinds of modes that were real-time adaptive control, fixed timing, and noncable coordination control, the system could set seventy green-wave lines under exceptional circumstances.

In theoretical research, Sen-Fa et al. (1998), Southeast University, proposed hierarchical structure of urban traffic liner control system. On the first level, the cycle and green ratio of traffic signal of a single junction was determined by fuzzy logic controller. On the second level, the phase offset between two adjacent junctions was obtained by fuzzy phase controller. And between two levels there was a fuzzy switch to coordinate them. So as to achieve the purpose of reducing vehicle delay, the author presented a specific implementation method which could effectively reduce the queue length [13]. Yun-Tao et al., Tongji University (2003), proposed a design method which used to optimize phase offset of coordination control on urban main roads based on a genetic algorithm [14]. Chu et al. model the role of the street environment in how people cross roads in urban settings. Respondents were placed in real traffic conditions at the curbside of street blocks in the Tampa Bay area for a three-minute observation of the street environment [15]. Ye et al. develop a method to separate the impacts on pedestrian traffic among different factors. Focusing on an individual pedestrian traffic indicator—walking speed—three important human factors, that is, gender, age, and luggage-carrying, were selected to investigate their respective impacts on walking speed [16].

2. Building the System Model

Large numbers of pedestrians crossing streets have become a significant feature of China's urban roads as travel increases. To solve this problem, many pedestrian mid-block crossings are set between adjacent intersections for the convenience of pedestrians. While most of the domestic pedestrian mid-block crossings lack reasonable and effective control, arterial traffic subject to certain constraints and obstacles.

On a section between two intersections, if a pedestrian mid-block crossing is controlled by one or more fixed-time and actuated signal controllers, coordination control of signal timing between pedestrians and intersections can be taken into consideration, which could reduce delays of vehicles on pedestrian mid-block crossings and improve safety of crossing streets.

Building a traditional artery or network coordination control models needs a "distance-time" graph [6], which is an intuitive and effective tool for coordination control studies. However, the author considers that it is not enough to use a "distance-time" graph. Traffic flow movement graph, also known as "flow rate-time" graph of traffic flow on sections, reflects the signal coordination relationship of pedestrian mid-block crossings and intersections as well. This paper will build a signal coordination control system model of pedestrian mid-block street crossings and intersections under different traffic control circumstances and take four pedestrian mid-block street crossings as an example which are on sections of the Taiping North Road in Nanjing for proving the feasibility of this model.

2.1. Building the “Distance-Time-Flow Rate” Graph

Not only the “distance-time” graph but also the traffic flow movement graph which is “flow rate-time” graph of traffic flow on sections reflect the signal coordination relationship of pedestrian mid-block crossings and intersections. The changes of the traffic flow can be observed directly from the flow graph.

The “vehicle arrival rate-time” relational graph of the stop line sections is a basis to determine all the traffic flow movement parameters (queue length, delay time, stop rate). For this reason, the traffic model which the signal coordination between the pedestrian mid-block street crossing and intersection relies on must be established on the basis of prediction of “vehicle arrival rate-time” graph.

If we added “flow rate-time” graph which reflects the flow rate changing to “distance-time” graph, and building the “distance-time-flow rate” graph of traffic flow movements, this would help a lot in finding the reasonable vehicle passing bandwidth and coordinating the initiate program. Besides, it also reflects system process parameters such as stop rate directly.

The steps to build a “distance-time-flow rate” graph (see Figure 1) between two sections are as follows.

- (1) Define principal vertical axis as distance (line out the relative positions of stop lines on sections), auxiliary vertical axis as flow rate, and horizontal axis as time, and build a graphic formula.
- (2) Draw a traffic flow rate-time graph of an initial traffic flow which comes from upstream sections.
- (3) Take a predicted drive-in vehicle flow rate-time graph on downstream sections on the basis of certain functional relations.
- (4) Design the distribution program of signal timing.
- (5) Set the starting point of pedestrian cycle signal on mid-block street crossings according to the designed green time offset.
- (6) Draw the signal phase diagram of pedestrian mid-block street crossings according to the signal timing distribution design and green time offset vehicle.

From Figure 1, all the traffic flows that arrive on the stop line are determined by a discharging law of upstream intersections. This means that if the functional relation between vehicle drive-out flow rate and time is known, we can calculate the functional relation between vehicle arrival rate and time on the stop lines. Speed discreteness should be considered when calculating the vehicle arrival flow rate-time graph on the stop line of pedestrian mid-block crossings, which is on the basis of drive-out rate graph of departing vehicles from upstream intersections. This paper uses geometric distribution function-Robertson [2] method to calculate vehicle arrival rate in a time interval:

$$q_{\partial}(i+t) = Fq_0(i) + (1-F)q_{\partial}(i+t-1) \quad (2.1)$$

(see [17]). Revise (2.1),

$$q_{\partial}(j) = \sum_{i=1}^{j-t} q_0(i)F(1-F)^{j-t-i}, \quad (2.2)$$

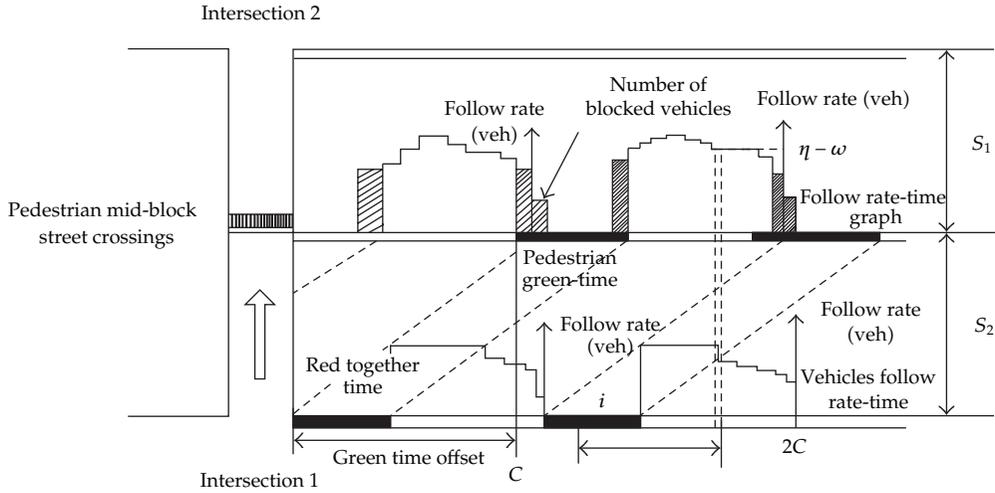


Figure 1: Distance-time-flow rate graph between two sections.

where $j = t + i$. $q_0(i + t)$ —predicted vehicle arrival rate on a downstream section at $(i + t)$ time interval; $q_0(i)$ —vehicle passing rate on the upstream stop line section at i time interval; t —0.8 times of vehicle average travel time which is between two sections above, using the quantity of time intervals as a unit; F —a factor which describes the extent of speed dispersion of traffic flow in the movement process, which is called traffic flow coefficient of dispersion, given by an empirical formula [18] as follows:

$$F = \frac{1}{1 + 0.35t}. \tag{2.3}$$

For a more convenient calculation, we divide a signal cycle into several smaller time periods. In these small time periods, it is supposed that the flow rate-time relation of vehicles arriving at intersections is of uniform distribution, which means to maintain the same value of flow rate which could be obtained by investigation. We can get a flow rate-time graph according to an average flow rate in a smaller period of time. And this could determine the vehicle flow rate which is from upstream intersections into road section of entry at the i time interval.

2.2. Calculate the Coordination Control of Pedestrian Mid-Block Street Crossing

To define the signal timing program of pedestrian mid-block street crossing needs to determine several controlling parameters, which are signal cycle length and green time of pedestrian mid-block crossings and phase offsets.

The cycle length and signal stages of pedestrian mid-block street crossings could be determined by “flow-rate-time” graph and constraint conditions of limit value of waiting time which pedestrian could tolerate. And these things above are confirmed by the predicted traffic flow that arrives at pedestrian mid-block street crossings.

For arterial roads and secondary arterial roads studied here, the road section traffic flow is of high volume and the vehicle speed is fast. As a result, the waiting time that pedestrian can tolerate is 50–60 seconds.

The flow rate of vehicles that enter the intersection is related to the phase design of an intersection signal cycle, so the predicted flow rate-time graph that arrives at the pedestrian mid-block street crossings is related with the phase of the upstream signal cycle. There are three trickles of vehicles entering the intersection from upstream road section, which are straight flow (S), right-entering right-turn flow (R), and left-entering left-turn flow (L). After merged, they are divided into five groups. Group 1 is (RSL) that 3 trickles of traffic flow are in the same phase, group 2 is ($R/S/L$) that 3 trickles of traffic flow are in different phases, group 3 is (RS/L) that R and S are in the same phase, group 4 is (SL/R) that S and L are in the same phase, and group 5 is (RL/S) that R and L are in the same phase.

When it is RSL , usually the signal cycle length of an intersection is short. As a result, the intersection cycle could be the pedestrian signal cycle C_p . So, there is a “unimodal” in the vehicle traffic flow rate in a cycle, which concludes that the green time of pedestrian mid-block street crossings g_p can be set by using the gap that is formed by “red together time” of the upstream signal. See Figure 2.

Under such circumstance, the pedestrian signal is the intersection signal cycle, and the green time of pedestrian mid-block street crossing is a function of cycle length. Hence, determining phase offset is necessary for the pedestrian signal distribution. We build a unary linear programming model to solve the optimal phase offset which is in (2.4). The planning intention is to make total delay of vehicles D_r in pedestrian green time minimum. The limit value of waiting time which pedestrians could tolerate is 60 seconds:

$$\min D_r = \sum_{j=(\text{OFS}+C_p-g_p)/n}^{(\text{OFS}+C_p)/n} nq_j(\text{OFS} + C_p - n_j), \quad (2.4)$$

$$\bar{t} - \frac{C_p}{2} \leq \text{OFS} \leq \frac{C_p}{2} + \bar{t}, \quad (2.5)$$

$$g_p = f(C_p),$$

$$0 \leq C_p - g_p \leq 60,$$

$$g_p = \frac{Q_{p \max}}{e_{pr} \times d} + \frac{l}{v_p} = \frac{(a_p + \alpha a_b + \beta a_h) \times C_p}{e_{pr} \times d} + \frac{l}{v_p} = f(C_p), \quad (2.6)$$

(see [6]), where D_r —delay of vehicles when it is green time for pedestrians and red time for vehicles (s); OFS —phase offset (s); C_p —cycle length of pedestrians, equaling the cycle length of intersections (s); g_p —green time of pedestrians (s); j —time interval in the green time; q_j —predicted flow rate of vehicles passing through stop lines of pedestrian mid-block street crossings which j corresponds with; n —length of time intervals (s); \bar{t} —average travel time which is from upstream intersections into stop lines of pedestrian mid-block street crossings on road sections, calculated by the average travel speed of road sections (s); $Q_{p \max}$ —number of traffic entities which is in a larger demand direction in a signal cycle, converted into units of persons (person); a_p , a_b , a_h —arrival rate of pedestrians, bicycles and help-move-vehicles separately in the large demanding direction (number of the arrivings/s); α , β —separate coefficient of conversion of bicycles, help-move-vehicles, and pedestrians ($\alpha = 5.6$, $\beta = 5.6$);

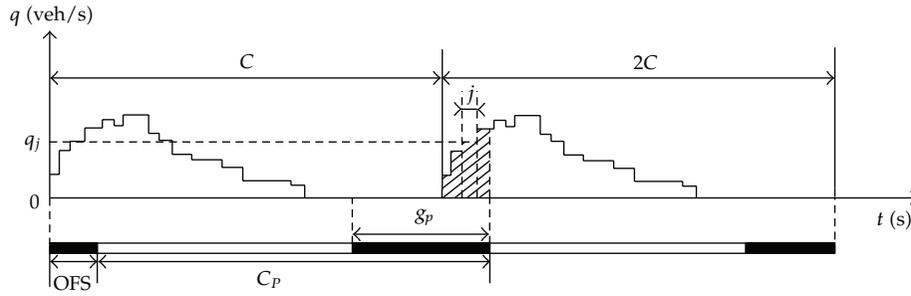


Figure 2: The establishment of first system optimal model on the basis of vehicle flow-time of crosswalk (signal single cycle of pedestrian).

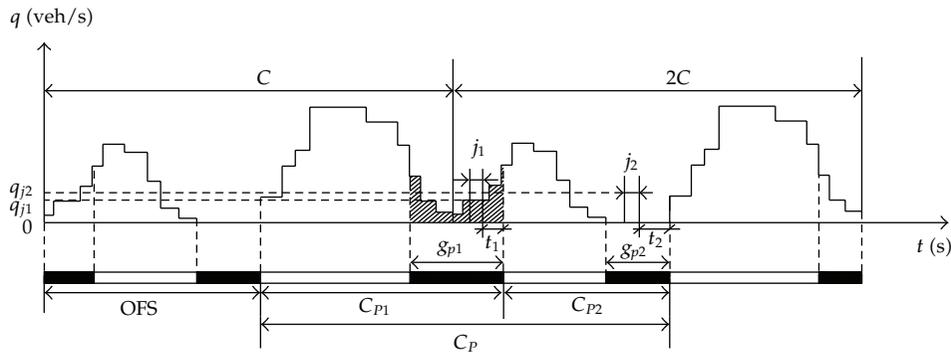


Figure 3: The establishment of first system optimal model on the basis of vehicle flow-time of crosswalk (signal double cycles of pedestrian).

e_{pr} —actual serving capacity of pedestrian mid-block crossings (person/green hour·m); d —width of pedestrian mid-block crossings (m); l —length of pedestrian mid-block crossings, (m); v_p —average speed of the pedestrian mid-block street crossing, taking it as 1.2 here (m/s); when it is $R/S/L$, the signal cycle is longer. There are “two humps” in a signal cycle. Besides, there is no gap that is formed by “red together time,” see Figures 3 and 4. In this circumstance, two departing opportunities can be given in one cycle for pedestrians. To set two green time of pedestrian mid-block street crossings g_1, g_2 , we can use the peak valley between “two humps.” This means to divide a cycle into two subcycles C_{P1}, C_{P2} , the sum of which equals the length of the intersection signal cycle. There is a pedestrian departing opportunity separately in each subcycle, and the length of each can be different.

On this occasion, the controls parameters are phase offset, length of subcycles and green time of the pedestrian signal. We could solve these parameters by solving a ternary linear programming model. Build the model like (2.7). This model is as a purpose of minimizing the vehicle delay D_r in the green time of pedestrians, and the constraint conditions are that the length of the intersection signal cycle is the sum of subcycles, and the sum of two waiting time of pedestrians is no more than the limit value (here is 60s).

the solution direction is correct, the computing time could be cut down. Usually the initial variable quantities are assigned as follows:

$$\text{OFS} = \frac{S_1}{v_d}, \quad C_{p1} = C_{p2} = \frac{1}{2}C_p, \quad (2.9)$$

where S_1 —distance from road sections to stop lines of pedestrian mid-block crossings that comes from upstream intersections (m); v_d —design speed on sections, taking the average speed value which has the highest frequency appearances; green flash time should also be set for warning the green time of pedestrians is going to finish. Usually it is 3–5 seconds.

3. Output of the System Model

3.1. Total Delay of Vehicles D

In the objective function of the system, D_r is the delay of vehicles when it is the pedestrian red light. But, in the beginning period of green time, the total delay of vehicles also contains D_s which is caused by the subsequent arriving vehicles blocked in a discharging red time. D_s is calculated by the delay time of vehicles [19] on the basis of the number of cumulative arriving of vehicles and the departing rate in green time. As a result, the total delay of vehicles is $D = D_r + D_s$.

If a cycle has two subcycles, calculating the total delays of subcycles, then the total vehicles' delays in a cycle are the sum of the total delays of subcycles.

3.2. Average Delay of Vehicles

The computing formula is

$$\bar{D} = \frac{D}{Q}, \quad (3.1)$$

where \bar{D} —average delay of vehicles (s/pcu); D —total delay of vehicles (s); Q —number of vehicles passing the stop lines in a cycle (pcu).

3.3. Stop Rate of Vehicles

In a cycle of vehicles driving on a section, stop rate of vehicles is a ratio of the number of vehicles which stop on the stop lines of the pedestrian mid-block crossings and the number of all vehicles passing the pedestrian crossing lines. The stopping rate of vehicles on a pedestrian crossing is calculated as follows:

$$E = \frac{\text{the number of vehicles stopping in a cycle}}{\text{the number of vehicles passing in a cycle}}. \quad (3.2)$$

3.4. Total Delay of Pedestrians

Delay of a single person involves waiting time and the reaction time of crossing green light. When calculating the total delays of pedestrian, we only calculate the sum of waiting time of pedestrians as the total delay of pedestrians D_p . According to the formula that calculates the crossing green time, we take the product [20] of the actual service ability and the width of pedestrian mid-block crossings as the dissipation rate, which is $e_{pr}d/3600$. The computing formula of unidirectional pedestrian crossing delay in a cycle is

$$D_{P_i} = \frac{e_{pr}da_i(C_p - g_p)^2}{2(e_{pr}d - 3600a_i)}, \quad (3.3)$$

where D_{P_i} —delay of unidirectional pedestrian mid-block crossings (s) ($i = 1, 2$); a_i —arrival rate of unidirectional pedestrians (person/s) ($i = 1, 2$).

For a pedestrian crossing, the total delays of pedestrian mid-block crossings are the sum of t bidirectional pedestrians, where $D_p = D_{P1} + D_{P2}$.

3.5. Average Delay of Pedestrians

After calculating the total delay of pedestrians, we know the quantity of arrival pedestrians in a cycle. The average delay of pedestrians can be calculated by using the following formula:

$$\bar{D}_p = \frac{D_p}{Q_p} = \frac{D_{P1} + D_{P2}}{C_p(a_1 + a_2)}, \quad (3.4)$$

a_i —arrival rate of unidirectional pedestrians, (person/s), ($i = 1, 2$).

When there are several pedestrian mid-block crossings, the total delay of vehicles, the average delay of vehicles, the stop rate of vehicles, and the total delay of pedestrians all take the sum of delays of the pedestrian mid-block street crossings while the average delays of pedestrians are the geometry average value of all the average delays of pedestrian mid-block street crossings.

4. Example Study

4.1. Investigation and Analysis of Taiping North Road

4.1.1. The Actuality of the Road Layout

Taiping Road is a main road of Nanjing. The road layout diagram of Taiping Road (Zhujiang-Beijing East Road) actuality is shown in Figure 5. This section has three T-shaped intersections which include an actuated pedestrian mid-block street crossing, a pedestrian overpass, and an uncontrolled pedestrian midblock street crossing.

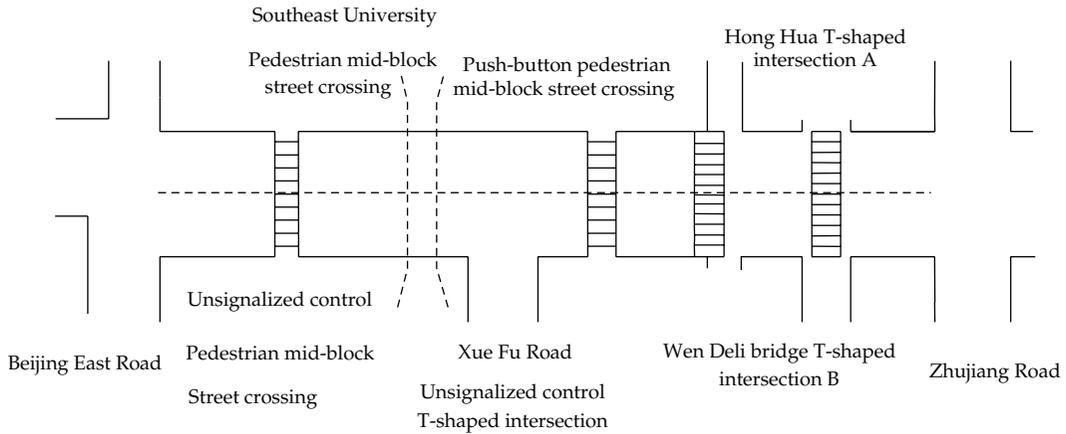


Figure 5: The road layout diagram of Taiping Road (Zhujiang-Beijing East Road) actuality.

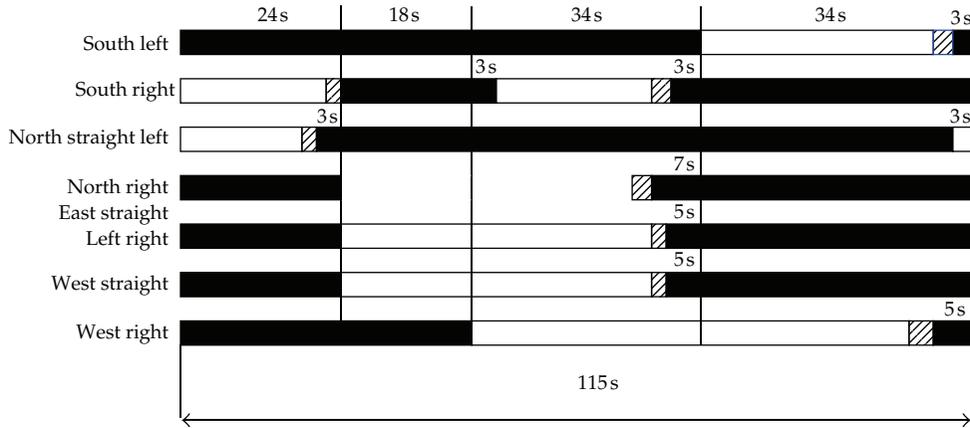


Figure 6: Distribution diagram of signal timing of Taiping North Road-Beijing East Road intersection.

4.1.2. Distribution of Signal Timing and Road Information

The timing signal controlled intersections on the section are Taiping North Road-Zhujiang Road intersection, Taiping North Road-Beijing East Road intersection, T-shaped intersection A, and T-shaped intersection B. The distribution sketch maps of signal timing are shown in Figures 6, 7, 8, and 9.

4.1.3. Traffic Characteristics of Vehicle

(1) Vehicle Speed of Road Section

Through the investigation of vehicle speed of Taiping North Road section, the highest frequency of the speed value group is 42.5 km/h.

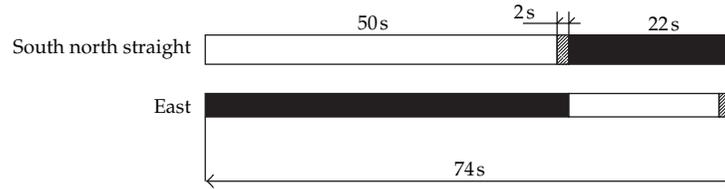


Figure 7: Distribution diagram of signal timing of T-shaped intersection A.

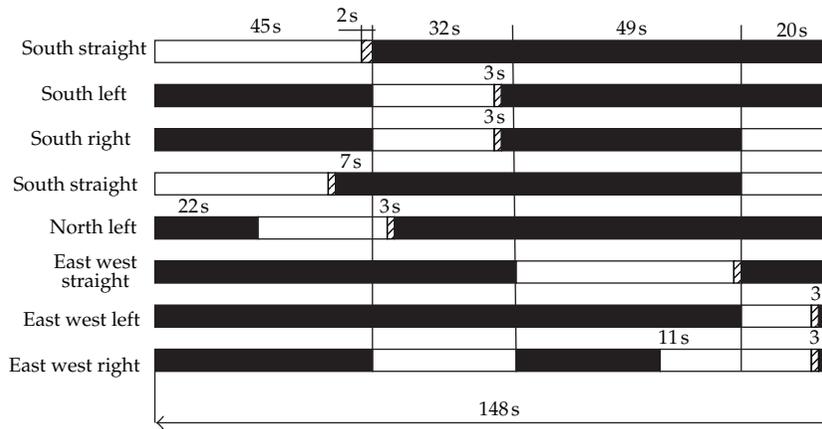


Figure 8: Distribution diagram of signal timing of Taiping North Road-Zhujiang Road intersection.

(2) Vehicle Delay of Road Section

Vehicle delay of road section is mainly produced in T-shaped intersection A and B and the actuated crosswalk C. The point investigation method is adopted to investigate vehicle delay at the peak time from the south to the north on the three intersections.

From the investigation and analysis of the three delays on Taiping Road section in the early peak time, the total delays of vehicles from south to north are 29673 s, the average delay is 24.79 s/pcu, and the stop rates before the stop line separately are 0.375, 0.34, and 0.45.

4.1.4. Traffic Characteristics of Mid-Block Street Crossing

(1) Arrival Characteristics of Mid-Block Street Crossing

Through surveys, the average traffic arrival rates of actuated crosswalk C and T-shaped junctions A and B are shown in Table 1.

(2) Traffic Speed of Mid-Block Street Crossing

Through investigation, the average mid-block street crossing speeds of pedestrians, bicycles and scooters are 1.2 m/s, 2.33 m/s, and 2.1 m/s on the road section.

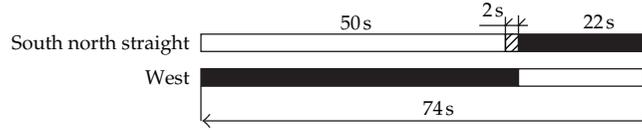


Figure 9: Distribution diagram of signal timing of T-shaped intersection B.

Table 1: The average traffic arrival rates of C, A, and B (person/s).

Direction	Crosswalk		
	Crosswalk C	T-shaped junction A	T-shaped junction B
From east to west	0.87	0.27	0.21
From west to east	0.12	0.18	0.20

(3) Traffic Delay of Mid-Block Street Crossing

Pedestrian delays of T-shaped junctions A, B, and C at the early peak time are calculated in Table 2.

4.2. Unidirectional Signal Coordination Control between Pedestrian and Intersection in Taiping North Road

A unidirectional signal coordination control between B, A, and C, three pedestrian mid-block street crossings, and the Taiping Road-Zhujiang Road intersection is realized on the road from south to north. First, B and upstream intersection are coordinated, then A and B are coordinated, finally A and the actuated crosswalk C are coordinated.

4.2.1. Signal Coordination Control of Wen Deli Crosswalk B and the Pacific North Road-Zhujiang Road

The signals of the upstream intersection into road section are merged in Figure 10.

Through investigation, the vehicle flow rate of the upstream intersection into road section is shown in calculation graph of section vehicle flow rate-time, taking 5 s as a time unit. The arrival flow rate-time graph of B stop line is acquired by (2.2).

The length of signal cycle of upstream intersection is 148 s, and pedestrian signal adopts the same signal cycle of the upstream intersection. Because the signal cycle time is too long, the signal of the pedestrian mid-block street crossing adopts dual cycle. From the data in Table 2 and the relationship of entering flow rate and time on B stop line and through the controls parameter optimization model formula (2.4) and (2.6), we can obtain that

$$\text{OFS is } 16 \text{ s}, \quad C_{p1} \text{ is } 73 \text{ s}, \quad C_{p2} \text{ is } 75 \text{ s}, \quad \min D_r \text{ is } 102 \text{ s/cycle}. \quad (4.1)$$

Then g_{p1} is 16 s and g_{p2} is 16 s. The 5 s at the end of two green lights is designed into green flash time.

If saturation flow rate of cross-section is taken as 0.84 pcu/s, then the total vehicle delay D is 263 pcu-s/cycle, that is 6397 pcu-s/h. The average delay of vehicles at B is $\bar{D} = D/Q = 4.90$, and stopping rate is 0.17.

Table 2: Pedestrian actuality delays of B, A, and C.

	B		A		C	
	From east to west	From west to east	From east to west	From west to east	From east to west	From west to east
Average delay (s/person-period)	24.5	24.4	25.3	24.5	20.4	33.6
Total delay (s/period)	416	395	553	357	1154	262
The total delay per hour of road section (s)	35600		40444		78425	



Figure 10: The merged signal of the upstream intersection into road section.

By (3.3) and (3.4), we can get that the average delay is 24.09 s/person from east to west, 25.2 s/person from west to east, and the total average delay is 24.6 s/person. Pedestrian total delay per hour of crosswalk B is 36413 s.

4.2.2. Signal Coordination Control between Hua Hongyuan Crosswalk A and Wen Deli Crosswalk B

Through investigation, the exiting flow rates of B intersection are shown in calculation graph of section vehicle flow rate-time. Take 5 s as a time unit. The entering flow rate-time graph on A stop line is gained via (2.2). From the data in Table 3, the entrancing flow rate-time relation of A stop line, and the controls parameter optimization model formula (2.6) and (2.7) iteration, we can get that OFS is 21 s, C_{p1} is 73 s, C_{p2} is 75 s, and $\min D_r$ is 8 s/cycle. Then g_{p1} is 16 s and g_{p1} is 16 s. The 5 s at the end of two green lights is designed into green flash time.

If saturation flow rate is taken as 0.75 pcu/s, then the total vehicle delay D is 277 pcu·s/cycle, that is 6757 pcu·s/h. The average delay of vehicles at A is $\bar{D} = D/Q = 5.17$ s/pcu and stop rate is 0.17.

By (3.3) and (3.4), we can get that the average delay is 24.7 s/person from east to west, 24.0 s/person from west to east, and the total average delay is 24.3 s/person. Pedestrian total delay per hour of crosswalk A is 39624 s.

4.2.3. Signal Coordination Control Design of Southeast University Actuated Crosswalk C and Hong Huayuan Crosswalk A

Through investigation, the exiting flow rates of A intersection are shown in calculation graph of section vehicle flow rate-time. Take 5 s as a time unit. The entering flow rate-time graph of C stop line is gained via (2.2). From the data in Table 3, the entering flow rate-time relation of

Table 3: Traffic characteristic of B, A, and C intersections.

Crossing	Parameter			Average arrival rate (person/s)
	l (m)	d (m)	v_p (m/s)	
B	18.6	6.5	1.2	0.21
A	18.6	5.6	1.2	0.27
C	18.6	5.6	1.2	0.87

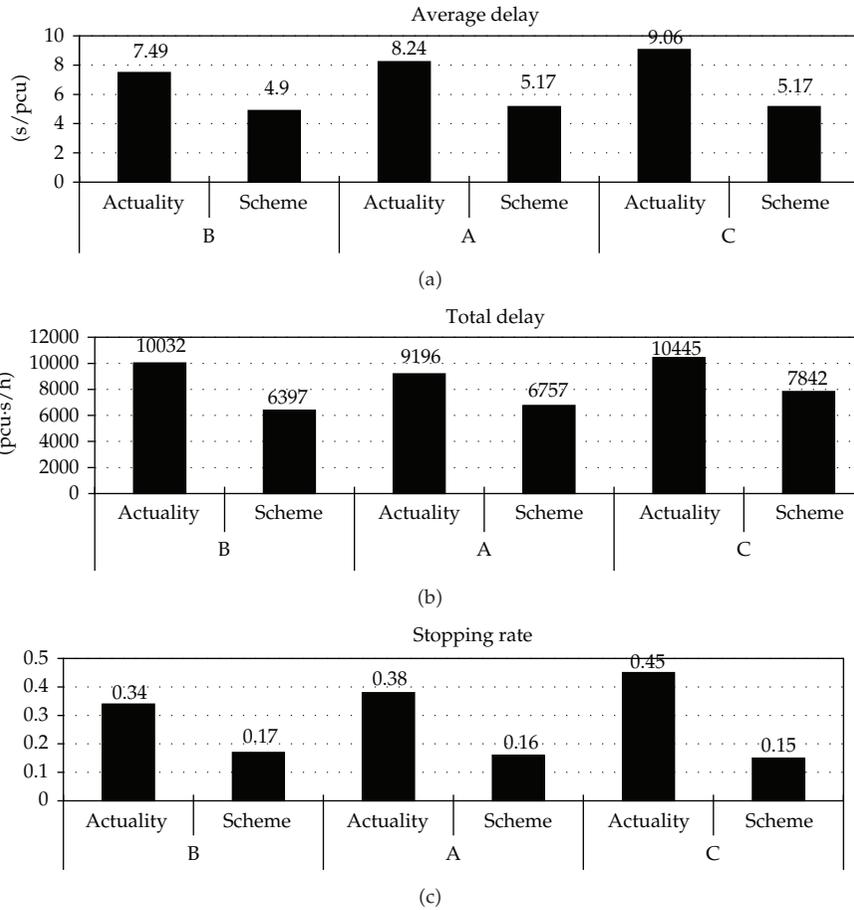
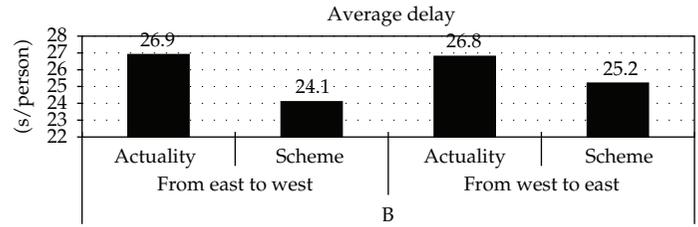


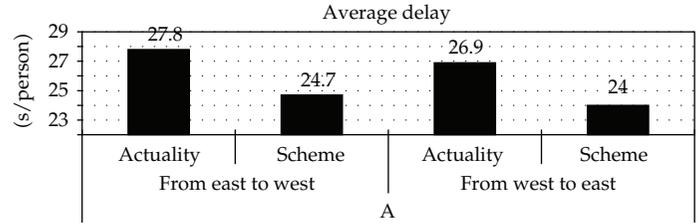
Figure 11: A comparison figure of vehicle delays from south to north of road actuality and scheme.

C stop line, and the controls parameter optimization model formula (2.6) and (2.7) iteration, we can get that OFS is 35 s, C_{p1} is 73 s, C_{p2} is 75 s, and $\min D_r$ is 24 s/cycle. Then t_{c1} is 16 s and t_{c2} is 16 s. The 5 s at the end of two green lights is designed into green flash time.

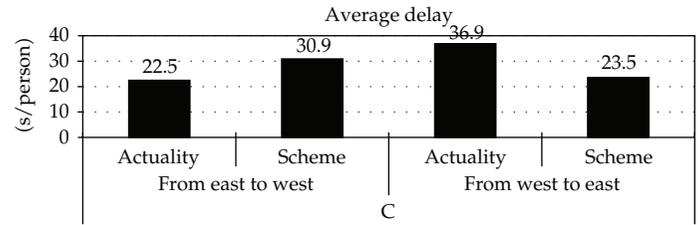
Through the delay calculating method of actuated signal vehicles and taking saturation flow rate as 0.75 pcu/s, the total vehicle delay D is 322 pcu-s/cycle, that is 7842 pcu-s/h. The average delay of vehicles at C is $\bar{D} = D/Q = 5.17$ s/pcu and stopping rate is 0.15.



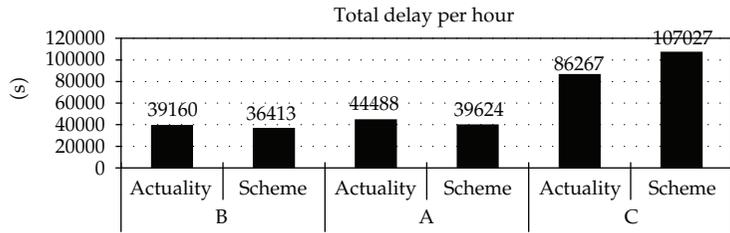
(a)



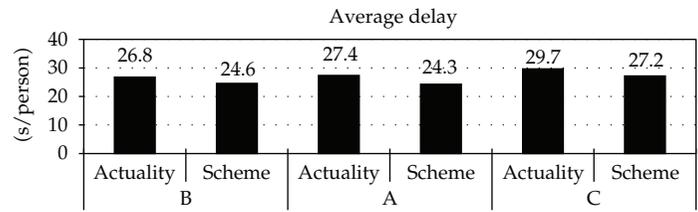
(b)



(c)



(d)



(e)

Figure 12: A comparison figure of pedestrian delays of road actuality and scheme.

By (3.3) and (3.4), we can get that the average delay is 30.9 s/person from east to west, 23.5 s/person from west to east, and the total average delay is 27.2 s/person. Pedestrian total delay per hour of crosswalk C is 107027 s.

4.2.4. Scheme Evaluation

According to the results of traffic actuality surveys and scheme output, a comparison table of vehicle delays and pedestrian delays from south to north of road actuality and scheme is gained and shown in Figures 11 and 12.

From Figures 11 and 12, it can be seen that vehicles' time benefits evaluation indexes which are vehicle delay and average delay of three crosswalks significantly decrease. And pedestrian mid-block street crossing time benefits evaluation indexes which are the average delay and total delay also decrease obviously after the scheme implemented. Time benefits of the scheme are obviously better than the current situation of signal scheme.

The time-efficient evaluation results of the vehicle from south to north are acquired only considering the vehicle unidirectional (from south to north) delay decreases. And the time-efficient evaluation of vehicles from north to south can also be acquired by the same method.

5. Conclusion

To coordinate the signal control between pedestrian mid-block street crossings and intersections, this paper proposes to use "distance-flow rate-time" graph as a tool to establish the system coordination model. The model for calculating delays of vehicles and passengers is established by using the method of coordinating the signal control between pedestrian mid-block street crossings and intersections.

Finally, taking one section on the Taiping North Road in Nanjing as an example, this paper carries out the coordination control of two timing-type crosswalks, a push-button (actuated) crosswalk, and the upstream intersection signal of Taiping Road and Zhujiang Road from south to north. Using the most optimal control plan, the system running parameters that are vehicles' delays, stop rate, and passengers' delays are obtained from the model. And though compared to current vehicles' delays at intersections from north to south, we conclude that the coordination control plan this paper proposes is effective.

Acknowledgments

The authors are grateful to the referee for the comments. This work is supported by the Jiangsu Province Natural Science Foundation (Bk2011745 of China, National key basic research and development program (973 program) issue (2012CB725402) of China).

References

- [1] W. Dian-Hai, *Traffic Flow Theory*, China Communications Press, Beijing, China, 2002.
- [2] M. Hosoi, S. Ishijima, and A. Kojima, "Dynamical model of a pedestrian in a crowd," in *Proceedings of the 5th IEEE International Workshop on Robot and Human Communication (RO-MAN)*, pp. 44–49, November 1996.
- [3] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research Part C*, vol. 19, no. 6, pp. 1202–1214, 2011.

- [4] 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.
- [5] M. S. Tarawneh, "Evaluation of pedestrian speed in Jordan with investigation of some contributing factors," *Journal of Safety Research*, vol. 32, no. 2, pp. 229–236, 2001.
- [6] Y. Pei-Kun and Z. Shu-Sheng, *Traffic Management and Control*, China Communications Press, Beijing, China, 2003.
- [7] *Criterion of Urban Transportation Planning (CJJ37-90)*, China Construction Industry Press, Beijing, China, 1991.
- [8] D. Sun, S. V. S. K. Ukkusuri, and R. F. Benekohal, "Modeling of motorist-pedestrian interaction at uncontrolled mid-block crosswalks," in *Proceedings of the TRB Annual Meeting*, 2003.
- [9] B. Wolshon and W. C. Taylor, "Analysis of intersection delay under real-time adaptive signal control," *Transportation Research Part C*, vol. 7, no. 1, pp. 53–72, 1999.
- [10] M. B. Trabia, M. S. Kaseko, and M. Ande, "A two-stage fuzzy logic controller for traffic signals," *Transportation Research Part C*, vol. 7, no. 6, pp. 353–367, 1999.
- [11] Y. Huang, Y. Kang, and S. Zhao, "Traffic signal coordination control method for river-crossing road bridge under load-bearing restriction," in *Proceedings of the International Conference on Transportation, Mechanical, and Electrical Engineering*, pp. 601–604, 2011.
- [12] L. Canqi and Y. Peikum, "Diffusion models of traffic platoon on signal-intersection and control of coordinated signals," *Journal of Tongji University*, vol. 24, no. 6, pp. 635–641, 1996.
- [13] C. Sen-Fa, C. Hong, and X. Ji-Qian, "Two level fuzzy control and simulation for urban traffic single intersection," *Journal of System Simulation*, vol. 10, no. 2, pp. 35–40, 1998.
- [14] C. Yun-Tao and P. Guo-Xiong, "Urban arterial road coordinate control based on genetic algorithm," *Journal of Traffic and Transportation Engineering*, vol. 3, no. 2, pp. 106–112, 2003.
- [15] X. Chu, M. Guttenplan, and M. Baltes, "Why people cross where they do: the role of the street environment," in *Proceedings of the TRB Annual Meeting*, 2003.
- [16] J. Ye, X. Chen, and N. Jian, "Impact analysis of human factors on pedestrian traffic characteristics," *Fire Safety Journal*, vol. 52, pp. 46–54, 2012.
- [17] Q. Yong-shen, *Urban Traffic Control*, China Communications Press, Beijing, China, 1989.
- [18] Y. Xiao-Guang, C. Bai-Lei, and P. Guo-Xiong, "Study of the way of setting pedestrian's traffic control signal," *China Journal of Highway and Transport*, vol. 14, no. 1, pp. 73–80, 2001.
- [19] W. Wei and G. Xiu-Cheng, *Traffic Engineering*, Southeast University Press, Nanjing, China, 2000.
- [20] Z. Zu-Wu, *Modern Traffic in City*, China Communications Press, Beijing, China, 1998.

Research Article

Analysis on the Synergy Evolutionary Development of the Collecting, Distributing, and Transporting System of Railway Heavy Haul Transportation

Fenling Feng, Dan Lan, and Liuwen Yang

School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

Correspondence should be addressed to Fenling Feng, ffl0731@163.com

Received 2 October 2012; Accepted 12 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Fenling 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.

A synergy evolutionary model of the collecting, distributing, and transporting system of railway heavy haul transportation is built by introducing synergy-related concepts and applying synergy evolutionary theory. Then spline interpolation method, numerical differential five-point formula, and method of least squares are used to solve synergistic coefficient, while fourth-order Runge-kutta method and fourth-order Adams linear implicit formula method are used to solve coevolutionary curve of the system. Finally, the heavy load transportation of Daqin Railway is an example of the empirical analysis. The research result shows that the degree of order of the system and its three subsystems—collecting, transporting, and distributing—increases as the synergetic coefficient of the subsystems increases; otherwise, the degree of the order will decrease. It also shows that this model can better analyze the coevolutionary process of the heavy load collecting, distributing, and transporting system of Daqin Railway, with its rationality and applicability verified.

1. Introduction

The collecting, distributing, and transporting system has served as a platform or bond to connect multiple means of transportation. It is critical to organize integrated transportation whose rapid and efficient implement requires orderly synergy between its three subsystems—collecting, transporting, and distributing. Studying the coevolutionary development of collecting, distributing, and transporting system helps to detect weaknesses and take corresponding measures to improve the orderly synergy of the entire system; it also helps to reveal laws concerning its evolution, making it more orderly and efficient.

Theory research of the coevolutionary development has hardly been done from the system point of view. Collaborative development in the field of transportation is mostly

about the synergy between the various modes of transportation. Studies are mainly focused on the measures of collaborative development, while few studies have been done to reveal the evolution law. Hook and Replogle [1], by taking the example of China, Japan and India, studied the street space to configure and use from the aspect of motorization influencing public policy. Adler et al. [2] explored the use of cooperative, distributed multiagent system to improve dynamic routing and traffic management. Hackney and Marchal [3] considered the influence of traveler transportation options which have optional range, mode and orientation preference, coordination degree, and alternative tools and can mobilize resources and other factors, using the simulation tool, the study of travelers' self-organization of transferring routing behaviors. Jeong et al. [4] discussed service chain which exists in relatively centralized flow, and the radial network hub has been widely applied in European freight railroad system, which provided reference to China's bulk cargo fast collection and distribution network. Konings [5] studied the integrated center with multimodal transport, storage, integrated with cargo transport and distribution, and offering good reference for railway transportation system. Fenling and Dan [6] constructed a cellular automata model for Daqin Railway in China by combining the rail transit system and cellular automata mode based on the four-aspect fixed block system, and Fenling and Feiran [7] set the incentive goal for cargo distribution and transportation system and further built an incentive model for the cargo distribution and transportation system that involves one principal and multiple agents. Jun et al. [8] used the data envelopment analysis model to evaluate the synergetic development of city tease based on the input-output matrix and the index of sustainable development, providing an approach of evaluating the validity degree of "synergic" and "development" within the city traffic system and among subsystems. Yan and Bei-Hua [9] analyzed the structure and synergetic character of MMTS with systematic science method and constructs the synergy equation for MMTS.

The coevolutionary development of collecting, distributing, and transporting subsystems has mainly been strengthened in practice by taking certain measures. Huaihai [10] used a three-dimensional modeling technology to analyze the system static structure, control element, and function simulation of the collecting and distributing system. Considering the system demand and the function of coupling system, he proposed a coupling system functional structure optimization method. Then he summarized transfer modes and spatial distribution modes among collecting and distributing subsystems in the passenger transport hub.

The coevolution has been widely used in the resource, environment, economic, and ecological systems. Norgaard [11] described the relationship between the economy and the environment. Coevolution theory is the combination of ecology and complex scientific synergy. Coevolution is not only the "synergy," but also the "evolution." It is the evolution relationship of various mutual influenced factors. Costanza et al. [12] argued that the system dynamics for the study of natural and man-made system relationship has important sense, and a broad interdisciplinary system concept has been applied to a series of technology. Holling [13] studied economic, environmental, and social relations. Aihua et al. [14] analyzed the interaction among subsystems and factors of water resources, social economy, and ecoenvironment and studied the evolving mechanism, mode, order parameter, controlling parameter, and synergetic frame of the water resources—social economy-eco environment. Yuming and Yan [15] analyzed the space-time distribution of the coordination degrees in China's 31 provinces in 1995, 2000, and 2005 empirically based on 18 indicators of economic subsystem and environmental subsystem and through employing models of coordination degrees of economic growth and environment and method of weighted entropy value.

Shangyou et al. [16] expounded the water resources ecological economic complex system, which is the objective reality of the practical and theoretical evolutionary processes of water resources planning and management based on positive methodology and system theoretical analysis. Xiangdong et al. [17] used the synergetics theory as a foundation to establish the multiplexed system's degree of order models and the multiplexed system's coordination evolution model through the system analysis of water resources restraint. Qingsong and Wenxiu [18] constructed a kind of practical calculation of composite system coordination degree model based on synergetics. Bak and Chen [19] described the concept of self-organization and its characteristic. Bobrek and Sokovic [20] discussed the management of modern enterprises for the implementation of total quality management and the synergistic effect of the general system theory.

This paper will mainly discuss degree of order, using synergy evolutionary theory to analyze the developing process of coevolution of the collecting, distributing, and transporting system and reveal the synergy evolutionary laws between all subsystems and laws of the entire system.

2. Synergy Evolutionary Model of the Collecting, Distributing, and Transporting System of Railway Heavy Haul Transportation

Synergy evolutionary model of the collecting, distributing, and transporting system of heavy haul transportation mainly builds its synergetic development model on the basis of the degree of order of its three subsystems.

2.1. Analysis on the Degree of Order of the Collecting, Distributing, and Transporting System of Railway Heavy Haul Transportation

Synergetics holds that synergy refers to the mutual coordination and synchronization between components in the developing process of the system, the degree of which is also called degree of order, and the synergetic effect determines the evolutionary trends and approaches of the system when it moves from disorder to order.

Assume that $n_i(t)$ ($i = 1, 2, 3$) refers to the degree of order of collecting subsystem, transporting subsystem, and distributing subsystem at the time of t , respectively, and all the parameters in the developing process of the subsystems are $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$, where $m \geq 1$, $\beta_{ij} \leq x_{ij} \leq \alpha_{ij}$, $j \in [1, m]$. Here, α_{ij}, β_{ij} refer to the upper bound and lower bound of parameter x_{ij} obtained from the stability critical point of the system. Without loss of generality, assume the greater the value of $x_{is+1}, x_{is+2}, \dots, x_{im}$, the lower the degree of order of the subsystems n_i , otherwise the higher the degree of order of the subsystems, n_i . A common method used to solve the degree of order of subsystem parameters is

$$n_i(x_{ij}) = \begin{cases} \frac{x_{ij} - \beta_{ij}}{\alpha_{ij} - \beta_{ij}} & j \in [1, s] \\ \frac{\alpha_{ij} - x_{ij}}{\alpha_{ij} - \beta_{ij}} & j \in [s+1, m] \end{cases} \quad i = 1, 2, 3, \quad (2.1)$$

where both the maximum value and minimum value of the parameters are transformed into 1 and 0, or 0 and 1 in the solution process when some parameters have wide range of variation

and some narrow range of variation. In such case, the relative role of parameters with narrow range of variation is exaggerated in the calculation of synergetic degree. Therefore, this paper adopts the following formula to solve the parameter's degree of order:

$$n_i(x_{ij}) = \begin{cases} \frac{x_{ij}}{2\alpha_{ij} - \beta_{ij}} & j \in [1, s] \\ 1 - \frac{x_{ij}}{2\alpha_{ij} - \beta_{ij}} & j \in [s+1, m] \end{cases} \quad i = 1, 2, 3. \quad (2.2)$$

From formula (2.2) we know that $n_i(x_{ij}) \in [0, 1]$, which is the degree of order of parameter x_i . The total contribution of parameter x_i to the degree of order of the subsystems can be realized through the integration of $n_i(x_{ij})$. The overall performance of subsystems depends not only on the numerical value of all parameters but also on their combining forms, for the forms directly determine the law of "integration." In practice, we often use geometric average method and linear weighted summation method to solve the degree of order of the subsystems. Thus we can obtain n_i :

$$n_i = \sqrt[m]{\prod_{j=1}^m n_i(x_{ij})} \quad \text{or} \quad n_j = \sum_{j=1}^m \omega_j n_i(x_{ij}) \sum_{j=1}^m \omega_j = 1, \quad (2.3)$$

where λ_j is the weight of the j th component of parameter x_i . From formula (2.3) we know that the degree of order $n_i \in [0, 1]$.

2.2. Synergy Evolutionary Model of the Collecting, Distributing, and Transporting System of Railway Heavy Haul Transportation

In order to show the mutual synergetic effect between all the subsystems, synergetic coefficient σ_{id} ($i \neq d$, $i, d = 1, 2, 3$) is introduced to represent the synergetic effect imposed by subsystem i on the degree of order of subsystem d . That is to say, the increase in the degree of order of subsystem i will push a lift in the degree of order of subsystem d , with $0 < \sigma_{id} \leq 1$; parameter ξ_i represents the combined effect imposed by the other two subsystems on subsystem i . In another word, it refers to the lift in the degree of order of subsystem i brought about by the increase in the orderly synergetic degree of the other two subsystems, with $0 < \xi_i \leq 1$. Thus we can obtain the synergy evolutionary equation of each subsystem:

$$\begin{aligned} \frac{dn_1}{dt} &= n_1 \left(\delta_1 - \lambda_1 n_1 + \sigma_{21} n_2 + \sigma_{31} n_3 + \xi_1 \frac{\min(n_2, n_3)}{\max(n_2, n_3)} \right), \\ \frac{dn_2}{dt} &= n_2 \left(\delta_2 - \lambda_2 n_2 + \sigma_{12} n_1 + \sigma_{32} n_3 + \xi_2 \frac{\min(n_1, n_3)}{\max(n_1, n_3)} \right), \\ \frac{dn_3}{dt} &= n_3 \left(\delta_3 - \lambda_3 n_3 + \sigma_{13} n_1 + \sigma_{23} n_2 + \xi_3 \frac{\min(n_1, n_2)}{\max(n_1, n_2)} \right), \end{aligned} \quad (2.4)$$

where δ_i ($i = 1, 2, 3$) is freely development coefficient (free from the influence of other subsystems and constrained by factors like its own environment), which represents

the development degree of degree of order of the three subsystems in the condition that they are free from the influence of other subsystems and only constrained by factors like its own environment, with $0 < \delta_i \leq 1$. With increase in the degree of order of the subsystems and restriction imposed by external resources, it will be more difficult for it to lift. Therefore, parameter λ_i refers to retardation in the degree of order inflicted by the subsystems themselves, with $0 < \lambda_i \leq 1$; $\min(n_i, n_j) / \max(n_i, n_j)$ refers to the synergetic degree between two subsystems.

The degree of order of the collecting, distributing, and transporting system of heavy haul transportation will be affected by that of its three subsystems. Set that the degree of order of all subsystem parameters is n_i^{k-1} , $i = 1, 2, 3$ at a given time t_{k-1} . In terms of time t_k which is in the developing and evolving process of the entire system, if the degree of order of all subsystem parameters is n_i^k , $i = 1, 2, 3$, the synergy evolutionary equation of synergetic degree of the system, N , will be

$$\frac{dN}{dt} = \theta \sqrt[3]{\left| \prod_{i=1}^3 [n_i^k - n_i^{k-1}] \right|}, \quad (2.5)$$

where

$$\theta = \frac{\min_i [n_i^k - n_i^{k-1}]}{\left| \min_i [n_i^k - n_i^{k-1}] \right| + \varepsilon} \quad (\varepsilon \rightarrow 0^+), \quad i = 1, 2, 3, \quad (2.6)$$

where $\min_i [n_i^k - n_i^{k-1}]$ may be 0. Therefore, an infinitesimal, ε , which is greater than 0, is added to denominator to make it unequal to 0. The role of parameter θ is to make sure that synergetic degree of the system will be positive if and only if $n_i^k - n_i^{k-1} > 0$, for all $i \in [1, 3]$.

According to the analysis above, the synergy evolutionary model of the collecting, distributing, and transporting system of heavy haul transportation can be,

$$n_i(x_{ij}) = \begin{cases} \frac{x_{ij}}{2\alpha_{ij} - \beta_{ij}}, & j \in [1, s] \\ 1 - \frac{x_{ij}}{2\alpha_{ij} - \beta_{ij}}, & j \in [s+1, m] \end{cases} \quad i = 1, 2, 3,$$

$$n_i = \sum_{j=1}^m \omega_j n_i(x_{ij}), \quad \omega_j \geq 0, \quad \sum_{j=1}^m \omega_j = 1, \quad i = 1, 2, 3,$$

$$\frac{dn_1}{n_1 dt} = \delta_1 - \lambda_1 n_1 + \sigma_{21} n_2 + \sigma_{31} n_3 + \xi_1 \frac{\min(n_2, n_3)}{\max(n_2, n_3)},$$

$$\frac{dn_2}{n_2 dt} = \delta_2 - \lambda_2 n_2 + \sigma_{12} n_1 + \sigma_{32} n_3 + \xi_2 \frac{\min(n_1, n_3)}{\max(n_1, n_3)},$$

$$\frac{dn_3}{n_3 dt} = \delta_3 - \lambda_3 n_3 + \sigma_{13} n_1 + \sigma_{23} n_2 + \xi_3 \frac{\min(n_1, n_2)}{\max(n_1, n_2)},$$

$$0 < \sigma_{id} \leq 1, 0 < \xi_i \leq 1, i, d = 1, 2, 3, i \neq d, 0 < \delta_i \leq 1, 0 < \lambda_i \leq 1,$$

$$\frac{dN}{dt} = \theta \sqrt[3]{\left| \prod_{i=1}^3 [n_i^k - n_i^{k-1}] \right|},$$

$$\theta = \frac{\min_i [n_i^k - n_i^{k-1}]}{\left| \min_i [n_i^k - n_i^{k-1}] \right| + \varepsilon}, \quad i = 1, 2, 3, \varepsilon \rightarrow 0^+.$$

(2.7)

3. Solution to the Synergy Evolutionary Model of Collecting, Distributing, and Transporting System of Railway Heavy Haul Transportation

3.1. Determining the Weight of Each Subsystem Parameter ω_j^i

In the evolving and developing process of the railway freight system, the influence imposed by different parameters on the system is different. So we need to determine the weights of all parameters.

(1) Standardization

Since the measurement unit of each parameter is different, it will lead to great difference in the measured value of each parameter. So standardization of the original data should be carried out in the first place to make all parameters dimensionless.

Assume that y_{ij}^i refers to the data of the j th parameter in the i th subsystem in the t th year ($i = 1, 2, 3, t = 1, 2, \dots, k, j = 1, 2, \dots, m$). Set \bar{y}_j^i to be the sample average of the j th parameter in the i th subsystem, then

$$\bar{y}_j^i = \frac{1}{k} \sum_{t=1}^k y_{tj}^i. \quad (3.1)$$

Set R_j^i to be the sample variance of the j th parameter in the i th subsystem, then

$$S_j^i = \left[\frac{1}{k-1} \sum_{t=1}^k (y_{tj}^i - \bar{y}_j^i)^2 \right]^{1/2}, \quad (3.2)$$

thus the standardized data is

$$Y_{tj}^i = \frac{y_{tj}^i - \bar{y}_j^i}{S_j^i}. \quad (3.3)$$

(2) *Determining the Correlation of Coefficient between All Parameters*

Correlation of coefficient reflects the degree to which all parameters affect each other. The greater the absolute value of the correlation of coefficient, the higher the degree that indicators affect each other, otherwise the lower the degree that indicators affect each other. The higher the total degree of correlation between a parameter and others, the greater the influence it imposes on other parameters. So we should give it a relatively great weight; otherwise, a relatively small weight will be given. Set $r_{k_1 k_2}^i$ to be the correlation of coefficient between the k_1 th and k_2 th parameters in the i th subsystem, and then the coefficient matrix of each parameter in the i th subsystem is R^i

$$R^i = \begin{pmatrix} r_{11}^i & r_{12}^i & \cdots & r_{1m}^i \\ r_{21}^i & r_{22}^i & \cdots & r_{2m}^i \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1}^i & r_{m2}^i & \cdots & r_{mm}^i \end{pmatrix}, \quad (3.4)$$

where

$$r_{k_1 k_2}^i = \frac{\sum_{j=1}^m (Y_{jk_1}^i - \bar{Y}_{jk_1}^i)(Y_{jk_2}^i - \bar{Y}_{jk_2}^i)}{\sqrt{\sum_{j=1}^m (Y_{jk_1}^i - \bar{Y}_{jk_1}^i)^2} \sqrt{\sum_{j=1}^m (Y_{jk_2}^i - \bar{Y}_{jk_2}^i)^2}} = \frac{\sum_{j=1}^m Y_{jk_1}^i Y_{jk_2}^i}{\sqrt{\sum_{j=1}^m (Y_{jk_1}^i)^2} \sqrt{\sum_{j=1}^m (Y_{jk_2}^i)^2}}, \quad (3.5)$$

when $k_1 = k_2$, $r_{k_1 k_2}^i = 1$.

(3) *Determining the Weight of Parameters*

Set

$$W_{k_2}^i = \sum_{k_1=1}^m |r_{k_1 k_2}^i| - 1, \quad (3.6)$$

then $W_{k_2}^i$ refers to the total degree of influence imposed by the k_2 th parameter in the i th subsystem on the rest parameters. Those rest parameters are $m - 1$ in total. If $W_{k_2}^i$ is relatively great, it indicates that the degree of influence imposed by the k_2 th parameter in the i th subsystem on the rest parameters is relatively great and we should give it a relatively great weight; otherwise, a relatively small weight will be given. Therefore, we can obtain the weight of parameter k_2 by normalizing $W_{k_2}^i$:

$$\omega_{k_2}^i = \frac{W_{k_2}^i}{\sum_{k_2=1}^m W_{k_2}^i}. \quad (3.7)$$

3.2. Determining Natural Growth Rate δ_i

The common method used to solve natural growth rate is first, solve the growth rate over the previous years and then use average growth rate to approximate the natural growth rate. But in the case that historical data are relative few, this method will fail to reflect the growth trends in a desirable way. Since the integration of collecting, distributing, and transporting system of heavy haul transportation in our country has started relatively late, our historical data are relatively few and the average growth rate of degree of order falls short to reflect its developing trends in a desirable way. Thus, this paper improves the average growth rate method and adopts the five-point numerical differential formula to solve the natural growth rate. First, we should deal with the original data by dividing a year's time into ten equal parts and using cubic spline interpolation (short for spline interpolation) to expand the data. Take the time step as $h = 0.1$ year. Assume that the value at the initial point is x_0 , then the value of the l th point, x_l , can be expressed as

$$x_l = x_0 + lh. \quad (3.8)$$

Next, we should use the numerical differential five-point formula to calculate the derivative of every point (including the original point and differentiation point). Thus we can obtain the general derivative formula of every point:

$$\begin{aligned} f'(x_0) &= \frac{-25f(x_0) + 48f(x_1) - 36f(x_2) + 16f(x_3) - 3f(x_4)}{12h}, \\ f'(x_1) &= \frac{-3f(x_0) - 10f(x_1) + 18f(x_2) - 6f(x_3) + f(x_4)}{12h}, \\ f'(x_l) &= \frac{f(x_{l-2}) - 8f(x_{l-1}) + 8f(x_{l+1}) - f(x_{l+2})}{12h}, \quad l = 3, 4, \dots, n-2, \\ f'(x_{p-1}) &= \frac{-f(x_{p-4}) + 6f(x_{p-3}) - 18f(x_{p-2}) + 10f(x_{p-1}) + 3f(x_p)}{12h}, \\ f'(x_p) &= \frac{3f(x_{p-4}) - 16f(x_{p-3}) + 36f(x_{p-2}) - 48f(x_{p-1}) + 25f(x_p)}{12h}. \end{aligned} \quad (3.9)$$

$f(x_l)$ refers to the cubic spline interpolation of x_l , $f'(x_l)$ the derivative of x_l , and p the number of points after data expansion. The precision of numerical differential five-point formula is $O(h^5)$, so in the collecting, distributing, and transporting subsystem we can obtain

$$\frac{dn_{il}}{n_{il}dt} = \frac{dn_{il}/dt}{n_{il}} = \frac{f'(n_{il})}{n_{il}}, \quad i = 1, 2, 3, \quad (3.10)$$

where n_{1l}, n_{2l}, n_{3l} are the points of degree of order over the previous years in the collecting, transporting and distributing subsystems obtained after using cubic spline interpolation, and formula (3.10) is used to calculate the growth rate of every point to solve the average growth rate and the natural growth rate of the three subsystems.

3.3. Solutions of the Self-Retardation Coefficients λ_i of Each Subsystem and the Synergistic Coefficients σ_{id} between These Subsystems

From formula (2.4) and formula (3.10), it comes to the following result:

$$\begin{aligned}\frac{f'(n_{1l})}{n_{1l}} &= \delta_1 - \lambda_1 n_{1l} + \sigma_{21} n_{2l} + \sigma_{31} n_{3l} + \xi_1 \frac{\min(n_{2l}, n_{3l})}{\max(n_{2l}, n_{3l})}, \\ \frac{f'(n_{2l})}{n_{2l}} &= \delta_2 - \lambda_2 n_{2l} + \sigma_{12} n_{1l} + \sigma_{32} n_{3l} + \xi_2 \frac{\min(n_{1l}, n_{3l})}{\max(n_{1l}, n_{3l})}, \\ \frac{f'(n_{3l})}{n_{3l}} &= \delta_3 - \lambda_3 n_{3l} + \sigma_{13} n_{1l} + \sigma_{23} n_{2l} + \xi_3 \frac{\min(n_{1l}, n_{2l})}{\max(n_{1l}, n_{2l})}.\end{aligned}\quad (3.11)$$

The left end of formula (3.11) is determined by formula (3.9). To solve the result, this paper applies the Method of Least Squares, which works out the parameter by finding the sum of squared errors between the minimized desired data and the real data. Therefore, formula (3.11) should firstly be transformed into the following formula:

$$\begin{aligned}S_1 &= \sum_{l=0}^p \left[\frac{f'(n_{1l})}{n_{1l}} - \left(\delta_1 - \lambda_1 n_{1l} + \sigma_{21} n_{2l} + \sigma_{31} n_{3l} + \xi_1 \frac{\min(n_{2l}, n_{3l})}{\max(n_{2l}, n_{3l})} \right) \right]^2, \\ S_2 &= \sum_{l=0}^p \left[\frac{f'(n_{2l})}{n_{2l}} - \left(\delta_2 - \lambda_2 n_{2l} + \sigma_{12} n_{1l} + \sigma_{32} n_{3l} + \xi_2 \frac{\min(n_{1l}, n_{3l})}{\max(n_{1l}, n_{3l})} \right) \right]^2, \\ S_3 &= \sum_{l=0}^p \left[\frac{f'(n_{3l})}{n_{3l}} - \left(\delta_3 - \lambda_3 n_{3l} + \sigma_{13} n_{1l} + \sigma_{23} n_{2l} + \xi_3 \frac{\min(n_{1l}, n_{2l})}{\max(n_{1l}, n_{2l})} \right) \right]^2,\end{aligned}\quad (3.12)$$

where S_i ($i = 1, 2, 3$) denotes the sum of squared errors of the No. i subsystem at p interpolating point.

Then, according to the Method of Least Squares, formula (3.12) should be transformed as follows:

$$\begin{aligned}\min \quad & S(\delta_1, \delta_2, \delta_3, \lambda_1, \lambda_2, \lambda_3, \sigma_{21}, \sigma_{31}, \sigma_{12}, \sigma_{32}, \sigma_{13}, \sigma_{23}, \xi_1, \xi_2, \xi_3) = \sum_{i=1}^3 S_i \\ & 0 < \delta_i \leq 1 \\ & 0 < \lambda_i \leq 1 \\ \text{s.t.} \quad & 0 < \sigma_{id} \leq 1 \\ & 0 < \xi_i \leq 1 \\ & i = 1, 2, 3 \quad d = 1, 2, 3 \quad i \neq d.\end{aligned}\quad (3.13)$$

By solving formula (3.13) through nonlinear optimization function `fmincon` of MATLAB, the synergistic coefficients of each subsystem can be figured out.

3.4. Solutions of Synergy Evolution Equations: Collecting, Distributing, and Transporting System of Railway Heavy Haul Transportation and Its Subsystems

By plugging the synergistic coefficients of each subsystem into formula (2.4) and then solving it with Ruge-kutta method, the synergy evolution curves of each subsystem's ordering degree can be worked out.

By putting the results of formula (2.4) into formula (2.5), the rate of change of the synergistic degree of order can be worked out. Ruge-kutta method is applied to solve an equation whose variable is related to time, that is, the variable's derivative of time is a function of the variable and time. Since the synergy evolution equation of each subsystem is related to a function about degree of order and time, Ruge-kutta method can be applied to solve it. However, the synergy evolution equation of the collecting, distributing, and transporting system of heavy haul transportation is the rate of change of the annually synergistic degree of order, which is a concrete number, so Ruge-kutta method cannot be applied. Instead, this paper applies fourth-order Adams linear implicit formula method to work out the result. The formula is as follows:

$$y_{j+1} = y_j + \frac{h}{24}(9f_{j+1} + 19f_j - 5f_{j-1} + f_{j-2}). \quad (3.14)$$

The truncation error of the fourth-order Adams linear implicit formula is

$$R_{n+1} = -\frac{19}{720}h^5 y_n^{(5)} + O(h^6), \quad (3.15)$$

where

- (i) y_j is the No. j year's synergistic degree of order of the collecting, distributing, and transporting system of heavy haul transportation,
- (ii) f_j is the No. j year's rate of change of the synergistic ordering degree of the collecting, distributing, and transporting system of heavy haul transportation,
- (iii) h is the time step, here referring to one year.

4. Empirical Analysis

4.1. Overview of the Collecting, Distributing, and Transporting System of Daqin (Datong to Qin Huangdao) Railway Heavy Haul Transportation

Daqin Railway heavy load transportation involves more than 100 coal mining enterprises, 4 nationwide big power grids, 10 big steel companies, more than 6000 industrial and mining enterprises which are coal-consuming, and 3 coal ports. From the logistics perspective, the processes of production, transportation, and marketing can be taken into a comprehensive consideration. Therefore, Daqin heavy load transportation system is consisted of three subsystems—collecting system, distributing system, and Daqin Railway. It is a large equipment-interconnected, production-linked integrated system. Its subsystems and itself have very obvious characteristics.

By the optimization and integration of the supply of goods, most of the coal resources from the three regions of Shaanxi province, Shanxi province, and the west of Inner Mongolia are transported through Daqin Railway. The wagon flow of Daqin line consists of two parts, namely, the one from the attracting area and the one from collecting and transporting line. At the loading spot along Daqin Railway attracting area, there are 20,000-ton freight train, 10,000-ton freight train, and ordinary freight train. Achieving point-to-point scheduled transporting organization, the 20,000-ton and 10,000-ton freight trains directly run through Hudong Station, alleviating the organizing pressure of Hudong Station. Ordinary freight trains are combined into 10,000-ton freight train at technical station; 10,000-ton freight trains of Dazhun line (Datong to Zhun Ger) and Dabao line (Datong to Baotou) are organized into 20,000-ton freight train at Hudong Station.

(1) The Collecting System of Daqin Railway

The collecting system of Daqin Railway is mainly composed of 19 coal enterprises, 100 loading spots, several collecting lines such as Beitongpu line (Datong to Taiyuan), Dazhun line, Dabao line, Yungang line (Datong to the mining area of Gaoshan county), Kouquan line (Datong to the mining area of Kouquan county), and many kinds of fixed equipments and locomotive equipments.

The wagon flow north-running through Daxin Station along Beitongpu line is organized into 10,000-ton freight train at Daxin Station, and 10,000-ton freight trains organized at the loading spot go directly through Daxin Station. Along Daxin-Hanjialing line, Bazhuang Station, Huai ren Station, Jinshatan Station, and Beizhuozhuang Station are capable of handling 20,000-ton freight trains, while the rest of the stations can handle 10,000-ton freight trains. The wagon flow of Yungang branch, the hub of Datong, can be organized into 20,000-ton train, so is the wagon flow of Hanjialing Station. The wagon flow of Kouquan branch can be organized into 10,000-ton train. Dabao line and Dazhun line are capable of 10,000-ton freight train, which will be made into 20,000-ton freight train at Hudong Station. Wagon flows from Yangyuan Station and Zhulu Station are also 10,000-ton freight train.

(2) The Distributing System of Daqin Railway

The distributing system of Daqin Railway mainly consists of three parts, including coal-consuming companies, such as 5 nationwide power grids, 10 steel companies, more than 6000 industrial and mining enterprises, Panshan Power Plant, Zunhua Power Plant, 3 coal ports of Qinhuangdao Port, Tianjin Port, Tangshan Port and 2 distributing lines of Jigang line and Qiancao line.

The trains arriving at Niucun II, Caofeidian Port, and Jingtang Port are all 20,000-ton freight trains. According to the request of unloading equipment, the 20,000-ton freight train can be unloaded after being divided into two 10,000-ton freight trains or be directly unloaded. The wagon flow running to Qinhuangdao Dong Station is 10,000-ton freight train. After arriving at Xizhangzhuang Station (Houying Station), it will be divided into 5,000-ton freight train. 10,000-ton freight trains running through Zuipingshan Station and Zunhuabei Station will be divided into 5,000-ton train and then sent to power plant. The 10,000-ton freight train to Tianjin Port will be divided into 5,000-ton freight train at Jixianxi Station and then sent to the port. The 10,000-ton freight trains to Jingqin line and Beijing will be divided into 5,000-ton freight train at Dashizhuang Station and Chawu Station and then go on running.

Table 1: The previous data of the collecting, distributing, and transporting system.

Year	Market share of Daqin Railway's coal transportation (%)	Collecting system				Transporting system				Distributing system		
		Loading spots	10,000-ton loading spots	20,000-ton loading spots	10,000-ton freight train per day	20,000-ton freight train per day	Train density	Daily average volume (million ton)	Max. daily volume (million ton)	Unloading spots	Use ratio of port (%)	Use ratio of unloading station (%)
2006	92.10	91	34	11	69.4	6.4	72	69.5	79.1	3	78.70	48.87
2007	90.20	116	42	11	60.1	22.4	83.2	83.23	99.87	4	82.23	49.68
2008	90.10	118	42	12	55.6	29.8	85.8	93.2	105	4	94.07	60.35
2009	90.20	129	43	14	40.9	36.5	90	90.5	111.35	5	83.43	48.81
2010	90.67	147	44	16	34.8	48.7	88.3	110.97	122.03	5	97.96	70.21
2011	90.98	147	44	16	35.5	59.6	97	120.6	129.9	5	114.49	82.48

4.2. Synergy Evolutionary Analysis of the Collecting, Distributing, and Transporting System of Daqin Railway Heavy Haul Transportation

From the beginning of 2006, the rail volume of Daqin Railway has basically kept a 50 million annual growth and reached 440 million in 2011. Therefore, this paper tries to analyze the synergy evolution process and tendency of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation by applying the data starting from 2006.

4.2.1. Determining the Weight of Each Parameter

This paper selects the main parameters of the three subsystems, respectively, and the parameters can be seen in Table 1. By plugging the data of each subsystem's parameters into (3.1)–(3.7), the weight of each subsystem's parameter can be worked out, as can be seen in Table 2.

4.2.2. Determining the Degree of Order of Each Subsystem

By plugging the data of Table 1 in (2.2), the degree of order for each subsystem can be worked out, as seen in Table 3 and Figure 1.

Table 2: Weight of each subsystem's parameters.

Market share of Daqin Railway's coal transportation	Collecting system			Transporting system			Distributing system					
	Loading spots	10,000-ton loading spots	20,000-ton loading spots	10,000-ton freight train per day	20,000-ton freight train per day	Train density	Daily average volume	Max daily volume	Unloading spots	Use ratio of port	Use ratio of unloading station	
	0.1728	0.2955	0.3081	0.2236	0.1956	0.2052	0.1956	0.1984	0.2052	0.2925	0.3396	0.3679

Table 3: The degree of order of each subsystem.

Year	Collecting subsystem	Transporting subsystem	Distributing subsystem
2006	0.6127	0.4277	0.4761
2007	0.6913	0.4967	0.5593
2008	0.7046	0.5233	0.5987
2009	0.7478	0.5186	0.6347
2010	0.8019	0.5624	0.6831
2011	0.8025	0.6176	0.7381

From Table 2 and Figure 1, we can see the overall degree of order of the three subsystems presenting a rising trend. Collecting system has the highest degree of order due to its fine integration of goods, and, with the construction of strategic loading spots, its efficiency will be much higher. But influenced by Hudong Station, its degree of order in 2011 changed a little. The degree of order of the distributing system ranks the second place because of its stable flow direction of goods and its increasing distributing capability. Consequently, its degree of order presents a relatively stable increase. The degree of order of transporting system stands the last place. Though the innovation of transporting organization has improved the volume of carriage, transporting system remains as the bottleneck for the synergistic development of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation.

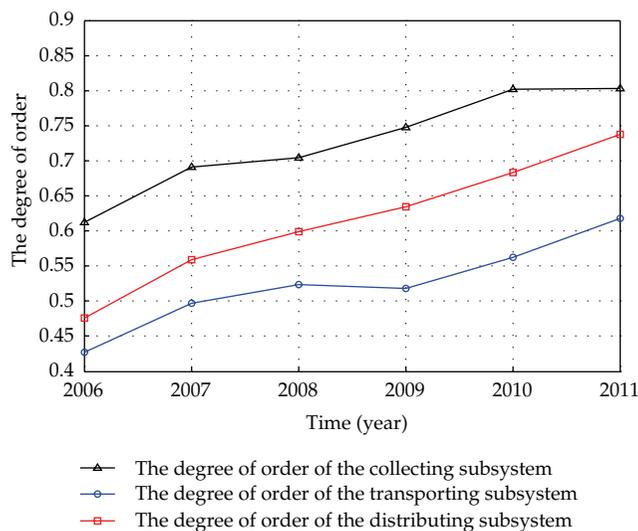


Figure 1: The change of degree of order of each subsystem.

4.2.3. Determining the Natural Growth Rate of Each Subsystem

By plugging the degree of order of the three subsystems into formulas (3.8)–(3.10), and applying to cubic spline interpolation and numerical differential five-point formula method, the natural growth rate of each subsystem can be figured out, respectively, as

$$\delta_1 = 0.0548, \quad \delta_2 = 0.0746, \quad \delta_3 = 0.0892. \quad (4.1)$$

The biggest natural growth rate belongs to distributing subsystem, which means that this subsystem enjoys the fastest development when only restricted by its own environment. Transporting subsystem develops relatively slower and collecting subsystem the slowest.

4.2.4. Solving the Self-Retardation Coefficients of all Subsystems and the Synergistic Coefficients between Them

By plugging the already solved natural rate of growth and the degree of order into formula (3.13), the self-retardation coefficients λ_i and the synergistic coefficients σ_{ij} and ξ_i can be worked out.

In collecting subsystem, $\lambda_1 = 0.4401$, $\sigma_{21} = 0.0112$, $\sigma_{31} = 0.0121$, and $\xi_1 = 0.3621$.

In transporting subsystem, $\lambda_2 = 0.6801$, $\sigma_{12} = 0.4199$, $\sigma_{32} = 0.0110$, and $\xi_2 = 0.0501$.

In distributing subsystem, $\lambda_3 = 0.8999$, $\sigma_{13} = 0.5999$, $\sigma_{23} = 0.0601$, and $\xi_3 = 0.1169$.

In terms of self-retardation coefficient we have the following.

- (1) The self-retardation coefficient of distributing subsystem is the biggest, which means that its degree of order gains the best development. Under the restriction of the available resource, it will be more and more difficult for its degree of order to rise.

- (2) The self-retardation coefficient of the collecting subsystem is the smallest, which indicates that there is still room for the improvement of its degree of order. With the stable implementation of the strategic loading spots and the improvement of loading equipments, its degree of order will rise.
- (3) The transporting subsystem's degree of order gains mild development compared to those of the other two subsystems' and still has room to be increased.

In terms of synergistic coefficient we have the following.

- (1) Due to the fact that collecting is the beginning part of the collecting, distributing, and transporting system and that its degree of order has a relatively large impact on the whole system, the collecting subsystem has the largest synergistic effect on the distributing and transporting subsystems; the synergistic degree between distributing and transporting, on the other hand, influences the degree of order of the collecting subsystem to the largest extent. For the reason that the collecting of goods and vehicles demands the cooperation between distributing and transporting so that the delivery of goods can be achieved, the synergistic degree between distributing and transporting subsystems has a relatively large impact on the collecting subsystem.
- (2) Distributing subsystem has the least synergistic effect on the other two in that the unloading points and the ports of the distributing subsystem have a capacity large enough to satisfy the demand of the collecting subsystem. Therefore, the distributing subsystem has the least influence on the degree of order of the other two, while the synergistic degree between "collecting" and transporting subsystems has an intermediate effect on the degree of order of the distributing subsystem.
- (3) The influence of the transporting subsystem on the other two is intermediate in that transporting connects collecting with distributing, and the self-retardation coefficient of the transporting subsystem itself is intermediate. Therefore, the transporting subsystem does not influence the other two on a large scale and the synergistic degree between the other two has the least influence on the degree of order of the transporting subsystem.

4.2.5. The Collecting, Distributing, and Transporting System of Daqin Railway Heavy Haul Transportation and the Solution of Equation of the Synergy Evolution of the Three Subsystems

The synergy evolution curves of the degree of order for the three subsystems can be drawn, respectively, through plugging the coefficients computed into (2.4) and solving the equation by the fourth-order Ruge-kutta method, to be seen in Figures 2, 3, and 4.

According to the synergistic theory, the cross-coupling of the input and the output, together with the nonlinear interaction among the three subsystems will drive the whole system to develop in a more organized and more efficient way until a relative stable balance is achieved, at which point the synergistic degree among the three subsystems will reach a maximum value. If there are no major events or special situations to interrupt, the stable state will last for a relatively long time until another chaos state presents itself.

From Figures 2, 3, and 4, it can be seen that the time for the three subsystems to reach a balance will be at about 2030 where the degrees of order for collecting, distributing, and

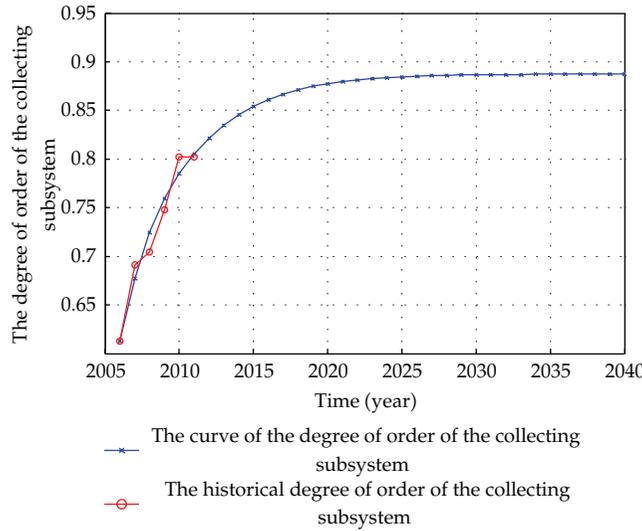


Figure 2: The synergy evolution curve of the degree of order of the collecting subsystem.

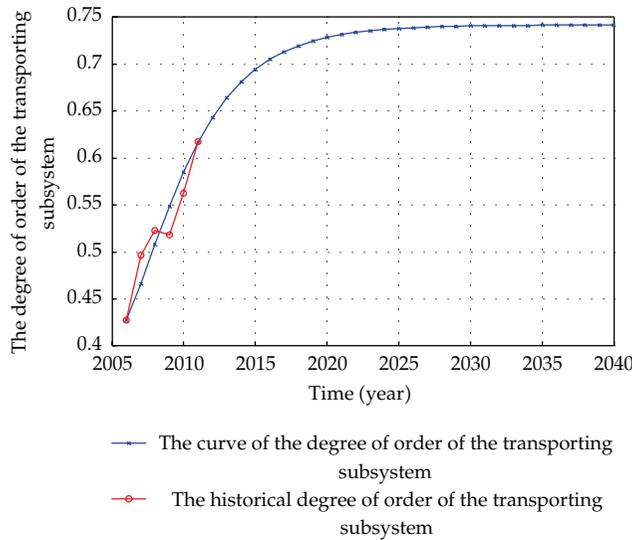


Figure 3: The synergy evolution curve of the degree of order of the transporting subsystem.

transporting will be 0.8870, 0.7481, and 0.8488, respectively, which, considering the resources available for the three subsystems, are the values that can be achieved in the condition of the largest degree of synergy. It can be seen, through matching the degrees of order of the three subsystems in the past years with the theoretical evolution curves of the three, that the results of the modeling are, in general, in line with the reality.

With 2011 as the base year, the absolute values for the growth of degree of order of the collecting, distributing, and transporting are 0.0845, 0.1305, and 0.1107, respectively. The rate of growth for the collecting subsystem is the smallest, for the order of degree of it is the largest among those of the three, which has reached 0.8025 in 2011. At such

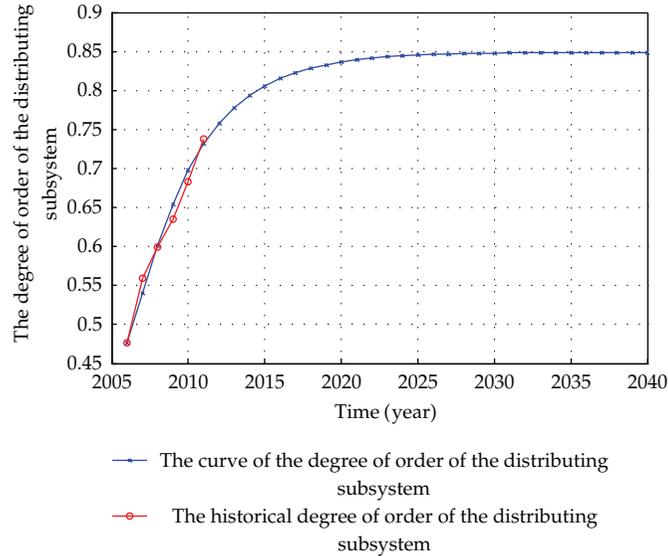


Figure 4: The synergy evolution curve of the degree of order of the distributing subsystem.

a high level, the growth of the degree of order will face much difficulty, hence a low rate of growth. The growth of the transporting subsystem is the largest due to two reasons: the degree of order of it is the smallest among the three, which restricts the synergistic development of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation, but will also increase greatly with the implementation of a series of transformation methods to expand capacity and the innovation of the transportation organizations; collecting subsystem has a large impact on the synergistic coefficient of the transporting subsystem, which means that once the order of degree of the collecting subsystem increases by 0.1, that of the transporting subsystem will increase by 0.04199, and, as a result of the synergistic promotion among the three, the degree of order of the transporting subsystem has the largest increase. The increase of degree of order of the distributing subsystem is intermediate but way above that of the collecting subsystem, for the former was 0.7381 in 2011 and stood at the second place among the three and still had room for growth. Besides, once the order of degree of the collecting subsystem increases by 0.1, that of the distributing will increase by 0.5999, and it will also increase if the synergistic order of degree of the collecting and transporting subsystems increases by 0.1. However, influenced by the self-retardation coefficient, its rate of increase is below that of the collecting subsystem.

The synergy evolution curve of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation can be drawn through plugging the degrees of order of the three subsystems calculated in (2.4) into (2.5) and solving the equation by the fourth-order Adams linear implicit formula method, to be seen in Figure 5.

It can be seen from Figure 5 that the development of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation will achieve a balance in 2030, where the order of degree of the system will be 0.8120, which, considering the resources and techniques available, indicates the largest degree of synergy that can be reached among the three subsystems.

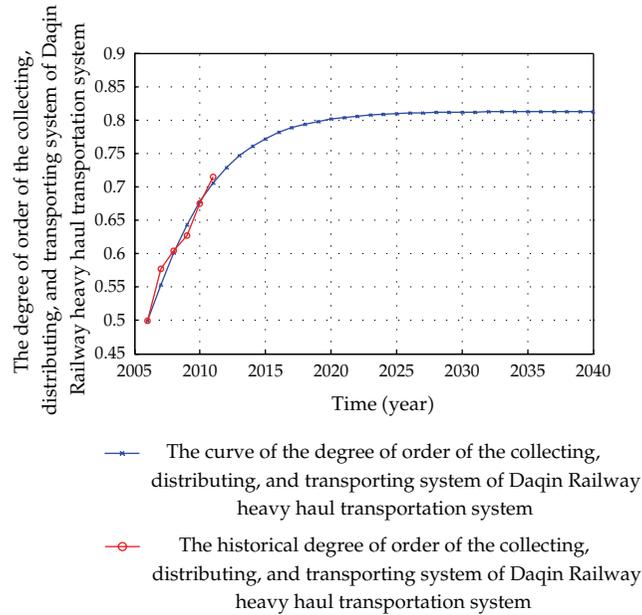


Figure 5: The synergy evolution curve of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation.

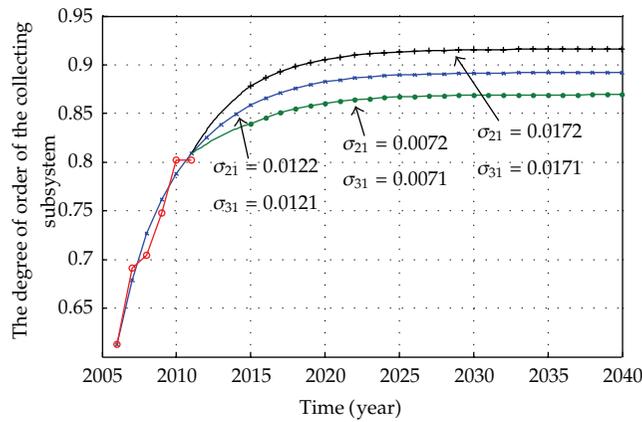


Figure 6: The ordered synergistic coevolution of the collecting subsystem with different synergistic coefficients.

4.2.6. The Analysis of the Synergy Evolutions of the Three Subsystems and the Collecting, Distributing, and Transporting System of Daqin Railway Heavy Haul Transportation

Based on the actual ordered synergistic coevolutions of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation and its subsystems, the ordered synergistic coevolutions of the system and its subsystems are analyzed, respectively, with a synergistic coefficient above and below the current value, to be seen in Figures 6, 7, 8, and 9.

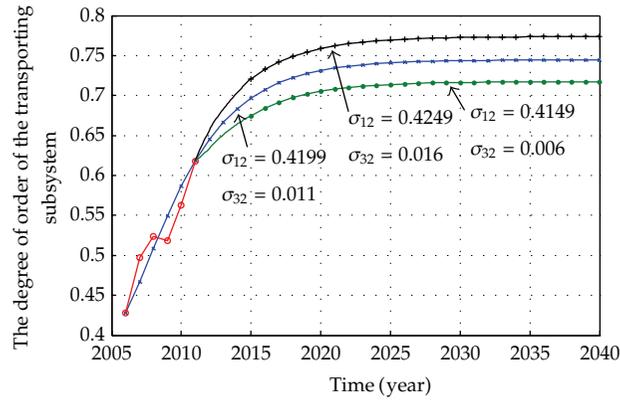


Figure 7: The ordered synergistic coevolution of the transporting subsystem with different synergistic coefficients.

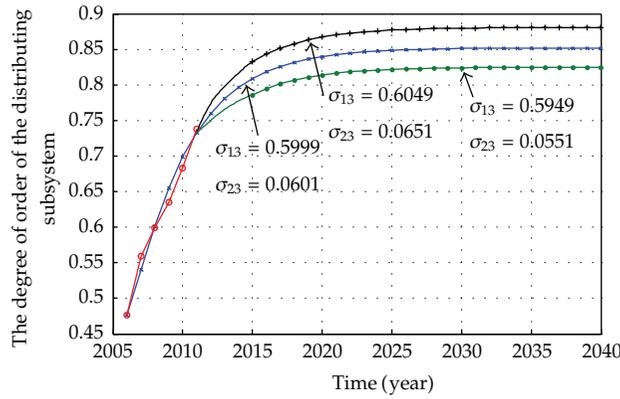


Figure 8: The ordered synergistic coevolution of the distributing subsystem with different synergistic coefficients.

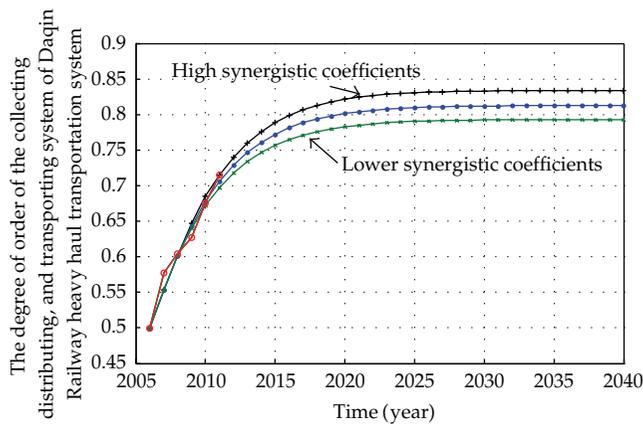


Figure 9: The ordered synergistic coevolution of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation with different synergistic coefficients.

Figure 6 reveals the ordered synergistic coevolution of the collecting subsystem with different synergistic coefficients, which when the synergistic coefficients the other two subsystems have on it, the σ_{21} and σ_{31} , are raised to 0.0172 and 0.0171, respectively, it will reach a balance in about 2030 and the degree of order at balance will be raised to 0.9163, while when the synergistic coefficients the other two subsystems have on it, the σ_{21} and σ_{31} , are lowered to 0.0122 and 0.0121, respectively, it will reach a balance in about 2030 and the degree of order at balance will be lowered to 0.8694.

Figure 7 reveals the ordered synergistic coevolution of the transporting subsystem with different synergistic coefficients, which when the synergistic coefficients the other two subsystems have on it, the σ_{12} and σ_{32} , are raised to 0.4249 and 0.0160, respectively, it will reach a balance in about 2030 and the degree of order at balance will be raised to 0.7736, while when the synergistic coefficients the other two subsystems have on it, the σ_{12} and σ_{32} , are lowered to 0.4149 and 0.0060, respectively, it will reach a balance in about 2030 and the degree of order at balance will be lowered to 0.7172.

Figure 8 reveals the ordered synergistic coevolution of the distributing subsystem with different synergistic coefficients, which when the synergistic coefficients the other two subsystems have on it, the σ_{13} and σ_{23} , are raised to 0.6049 and 0.0651, respectively, it will reach a balance in about 2030 and the degree of order at balance will be raised to 0.8860, while when the synergistic coefficients the other two subsystems have on it, the σ_{13} and σ_{23} , are lowered to 0.5949 and 0.0551, respectively, it will reach a balance in about 2030 and the degree of order at balance will be lowered to 0.8249.

Figure 9 reveals the ordered synergistic coevolution of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation with different synergistic coefficients, which when the ordered synergistic degree among the three subsystems is raised, it will reach a balance in about 2030 and the degree of order at balance will be raised to 0.8350, while when the ordered synergistic degree among the three subsystems is lowered, it will reach a balance in about 2030 and the degree of order at balance will be lowered to 0.7899.

It can be concluded from the above analyses that the ordered synergy within and among the three subsystems contributes to the uprising of the degree of order of the whole system and benefits the more organized and more efficient development of the collecting, distributing, and transporting system of Daqin Railway heavy haul transportation. Since the collecting, distributing, and transporting system of Daqin heavy haul transportation has only existed for a relatively short time, the parameter data may fluctuate.

5. Conclusion

(1) The paper builds a synergistic evolution model for the collecting, distributing, and transporting system of railway heavy haul transportation through the adoption of the concepts relevant to the degree of synergy of complex system and the application of the coevolution theory. The model is then transformed by spline interpolation method and five-point numerical differentiation formula into a least squares problem with linear constraint conditions, which is solved by the use of fourth-order Runge-kutta method and fourth-order Adams linear implicit formula method.

(2) The degrees of order of the collecting, distributing, and transporting system of railway heavy haul transportation and its three subsystems increase when the synergistic coefficient among the three subsystems increases and decrease when the latter decreases.

(3) The model used in the empirical analysis of the Daqin Railway can preferably analyze the ordered synergistic evolution of the collecting, distributing and transporting system of Daqin Railway heavy haul transportation. The collecting, distributing, and transporting system of Daqin Railway heavy haul transportation and the collecting, distributing, and transporting subsystems will reach a balance about 2030, where the order of degree of the system and the three subsystems will be 0.8120, 0.8879, 0.7481, and 0.8488, respectively, which, considering the resources available at present, are the values that can be achieved under the condition of the largest degree of synergy. Therefore, the model is proved to be reasonable and feasible.

Acknowledgment

This work was supported by the Fundamental Research Funds for the Central Universities (Grant no. 2010QZZD021).

References

- [1] W. Hook and M. Replogle, "Motorization and non-motorized transport in Asia: transport system evolution in China, Japan and Indonesia," *Land Use Policy*, vol. 13, no. 1, pp. 69–84, 1996.
- [2] J. L. Adler, G. Satapathy, V. Manikonda, B. Bowles, and V. J. Blue, "A multi-agent approach to cooperative traffic management and route guidance," *Transportation Research Part B*, vol. 39, no. 4, pp. 297–318, 2005.
- [3] J. Hackney and F. Marchal, "A coupled multi-agent microsimulation of social interactions and transportation behavior," *Transportation Research Part A*, vol. 45, no. 4, pp. 296–309, 2011.
- [4] S. J. Jeong, C. G. Lee, and J. H. Bookbinder, "The European freight railway system as a hub-and-spoke network," *Transportation Research Part A*, vol. 41, no. 6, pp. 523–536, 2007.
- [5] J. W. Konings, "Integrated centres for the transshipment, storage, collection and distribution of goods: a survey of the possibilities for a high-quality intermodal transport concept," *Transport Policy*, vol. 3, no. 1-2, pp. 3–11, 1996.
- [6] F. Fenling and L. Dan, "Heavy-haul train's operating ratio, speed and intensity relationship for Daqin railway based on cellular automata model," *Information Technology Journal*, vol. 11, no. 1, pp. 126–133, 2012.
- [7] F. Fenling and L. Feiran, "Railway heavy-haul cargo distribution and transportation system incentive mechanism based on principal-agent theory," *Information Technology Journal*, vol. 11, no. 7, pp. 884–890, 2012.
- [8] Z. Jun, D. Wen, and Z. Yue, "Integrated evaluation for synergetic development of city traffic system based on DEA model," *Computer Engineering and Applications*, vol. 43, no. 32, pp. 245–248, 2007.
- [9] Z. Yan and Z. Bei-Hua, "Evaluative equation for container multi-modal transport system," *Navigation of China*, no. 4, pp. 92–104, 2006.
- [10] T. Huaihai, "High-speed railway passenger hub for collection and distribution system structure coupling," Southeast University, Nanjing, China, 2009.
- [11] R. B. Norgaard, "Environmental economics: an evolutionary critique and a plea for pluralism," *Journal of Environmental Economics and Management*, vol. 12, no. 4, pp. 382–394, 1985.
- [12] R. Costanza, L. Wainger, and C. Folke, "Modeling complex ecological economic systems," *Bio Science*, vol. 43, no. 8, pp. 545–556, 1993.
- [13] C. S. Holling, "Understanding the complexity of economic, ecological, and social systems," *Ecosystems*, vol. 4, no. 5, pp. 390–405, 2001.
- [14] L. Aihua, L. Yuanyuan, and L. Jianqiang, "Preliminary discussion on synergetic development between water resources and social economy and eco-environment," *Yangtze River*, vol. 42, no. 18, pp. 117–121, 2011.
- [15] W. Yuming and Z. Yan, "Coupled coordinated development of regional economic growth and environmental research," *Resources Science*, vol. 30, no. 1, pp. 25–30, 2008.

- [16] F. Shangyou, L. Guoquan, and M. Yadong, "Water resources ecological and economic system and its sustainable development," *Engineering Journal of Wuhan University*, vol. 28, no. 6, pp. 624–629, 1995.
- [17] M. Xiangdong, S. Jinhua, and H. Zhenyun, "Co-evolutionary study on the ecological environment and the social economy multiplexed system," *Advances in Water Science*, vol. 20, no. 4, pp. 566–571, 2009.
- [18] M. Qingsong and H. Wenxiu, "Study of the coordinating measurement model with respect to composite system," *Journal of Tianjing University*, vol. 33, no. 4, pp. 444–446, 2000.
- [19] P. Bak and K. Chen, "Self-organized criticality," *Scientific American*, vol. 264, no. 1, pp. 26–33, 1991.
- [20] M. Bobrek and M. Sokovic, "Integration concept and synergetic effect in modern management," *Journal of Materials Processing Technology*, vol. 175, no. 1–3, pp. 33–39, 2006.

Research Article

A Study on the Coordination of Urban Traffic Control and Traffic Assignment

**ZhaoWei Qu, Yan Xing, XianMin Song,
YuZhou Duan, and Fulu Wei**

College of Transportation, Jilin University, Changchun 130025, China

Correspondence should be addressed to XianMin Song, 317146898@qq.com

Received 20 August 2012; Accepted 15 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 ZhaoWei Qu 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 interactions between signal setting and traffic assignment can directly affect the urban road network efficiency. In order to improve the coordination of signal setting with traffic assignment, this paper created a traffic control algorithm considering traffic assignment; meanwhile, the link impedance function and the route choice function were introduced into this paper to study the user's route choice and the road network flow distribution. Then based on the above research, we created a system utility value model. Finally through the VISSIM software to simulate the test network, we verified the superiority of the coordination algorithm and the model and gave the optimal flow of the road network.

1. Introduction

In urban transportation systems, the equilibrium of the road network supply and demand is an important problem, which can directly impact the road network efficiency. For a fixed road structure and traffic flow, signal setting is the primary influencing factor to traffic supply, and traffic assignment is the primary influencing factor to traffic demand [1, 2]. In order to balance the traffic demand and supply of urban road network and maximize the efficiency of transportation resources, experts and scholars from all over the world have done a lot of research and made great progress in the coordination of signal setting with traffic assignment aspects.

Firstly, Allsop [3] and Gartner [4] researched the relationship between signal setting and traffic assignment, and held that signal setting strategy can directly affect user's route choice and traffic flow distribution. Then they proposed a solution program to traffic equilibrium problem by alternately updating green split signal settings for fixed flows. On these bases, in order to achieve the global optimization, Tan et al. [5] and Smith [6, 7] converted

the signal setting and traffic assignment into a hybrid optimization problem. Their research showed that in some actual situation traffic setting even had deteriorated travel time by 30%; thus, they proposed a new signal setting policy under the constraint of stable traffic supply and traffic demand. Subsequently, Gartner et al. [8] and Fisk [9] proposed that we may take the signal setting and traffic assignment as a game problem of Stackberg, which regards traffic managers as superior decision-makers and travelers as subordinate decision-makers. Then, Yang and Yagar [10] proposed a bilevel programming model based on the above method; the model indicated that we can get queuing delay of saturated sections by the network equilibrium parameters instead of conventional delay calculation formula.

Since some traffic control systems and algorithms often require the real-time traffic flow distributed information to make traffic management strategy, Gartner et al. [11, 12] presented a combined framework about the real-time signal setting and traffic assignment, which can make appropriate dynamic control strategies and avoid traffic congestion according to dynamic traffic assignment model. With the research continued, Lee and Machemehl [13] studied the stochastic heuristic algorithms in coordination with signal setting and traffic assignment and applied the genetic algorithms in it. Chen and Ben-Akiva [14] regarded dynamic traffic control and traffic assignment as a noncooperation game problem; we could get the optimal signal setting and traffic assignment by iteration.

In recent years, many new methods emerged on the study of signal setting and traffic assignment relationship, for example, Chiou [15] introduced derivative information of the signal control variables into the TRANSYT system and then estimated the average vehicle delay in the road network based on sensitivity algorithm. Cipriani and Fusco [16] discussed the global optimization problem of signal setting and traffic assignment based on gradient projection method. Cascetta et al. [17] studied the optimization models and algorithms for urban signal settings with a small-scale real network based on stochastic traffic assignments. In the study of Gao et al. [18] and Chiou [19], the problem of signal setting and traffic assignment relationship was formulated as a nonlinear mathematical program with equilibrium constraints, in which the optimal signal settings were determined while trip rates and network flow are in equilibrium.

By summarizing the above research experience, this paper proposed a synergetic algorithm of dynamic signal setting and traffic assignment. In the algorithm, traffic control parameters can be updated in real time with the change of traffic demand. The basic idea of the algorithm is to change the road impedance by adjusting the split of the signal setting scheme and then cause the traffic volume transfer to realize the aim of the road network flow redistribution and the systems' efficiency optimization. Therefore, in order to further clarify the mechanism of signal setting and traffic assignment, this paper creates a simple and practical integrated model by introducing the link impedance function and the route choice function. Finally, a system utility value calculation method is given, by which we can get the optimal flow and optimal path impedance and then compare the real flow and path impedance with the optimal flow and optimal path impedance to determine whether the road network is in equilibrium. We use the above discrimination method to optimize the efficiency of the road network.

2. Problem Formulation

The coordination of signal setting and traffic assignment problem can be formulated as the problem of finding out the relationship between traffic demand and traffic supply, which is the problem of finding out the relationship between signal control parameters and user's

route choice. Firstly, we consider a traffic network $G(P, L)$ consisting of a set of nodes P and connected by a set of directed arcs L . Since traffic supply mainly includes traffic control and basic road information, the initial traffic control program setup is based on historical flows from road network, which adjusts the traffic flow of road network by signal parameters, such as signal cycle and green time. Thus, we can describe the relationship between the above items in the following equation:

$$U = f(C_j, g_j) = f_i(q_{k0}, c_{k0}), \quad (2.1)$$

where U is traffic signal control program; C_j is signal cycle of junction j ; g_j is green time of junction j ; q_{k0} is the initial traffic flow of link k ; c_{k0} is traffic capacity of link k ; i is the signal phase of junction.

Meanwhile, the user's route choice is usually affected by the signal control program, which is an important factor in restricting the road network flow distribution; another important restrictive factor is the origin-destination (O-D). Thus, we can describe the relationship between signal control program, user's route choice, and O-D in the following equation:

$$F_k = f(P_k^{rs}, U_i, L_i^{O,D}), \quad (2.2)$$

where F_k is the actual traffic flow of link k ; P_k^{rs} is the probability of path k being selected from origin (r) to destination (s); $L_i^{O,D}$ are the origin location and the destination location of traffic flow.

Since the effects of signal control program and traffic flow are reflected in traffic participant's travel time, we can analyze the traffic supply and traffic demand to determine the optimal flow of road network and then compare the difference between the actual travel time and the optimal travel time to discriminate whether the road network is in equilibrium.

On the basis of the above statement, two questions can be summed up from the above representations. First, a coordination algorithm of traffic demand and traffic supply should be established if you want to achieve the mutual feedback traffic control and flow distribution. Second, we should create some appropriate model about traffic control and traffic assignment to implement the algorithm.

3. Algorithm Establishments

The algorithm establishment consists of two parts in this paper: the first part of the algorithm mainly analyzes the relationship between signal setting and flow distribution. Before we establish the algorithm, a stable state of the road network flow is needed to realize the objective of traffic flow equilibrium, which means that the difference between road network inflow (q_{ui}) and road network outflow (q_{vi}) should be less than a fixed value (ε), $q_{ui} - q_{vi} < \varepsilon$, where, q_{ui} is the total vehicles enter the fixed road network; q_{vi} is the total vehicles leave the fixed road network. Meanwhile, the similarity degree of the path impedance needs to be known; we can describe it as follows: $t_k^{r-s} / t_{k+1}^{r-s} \approx 1$, where t_k^{r-s} is the impedance of path (k) from origin (r) to destination (s). The relationship between signal setting and flow distribution is shown in Figure 1.

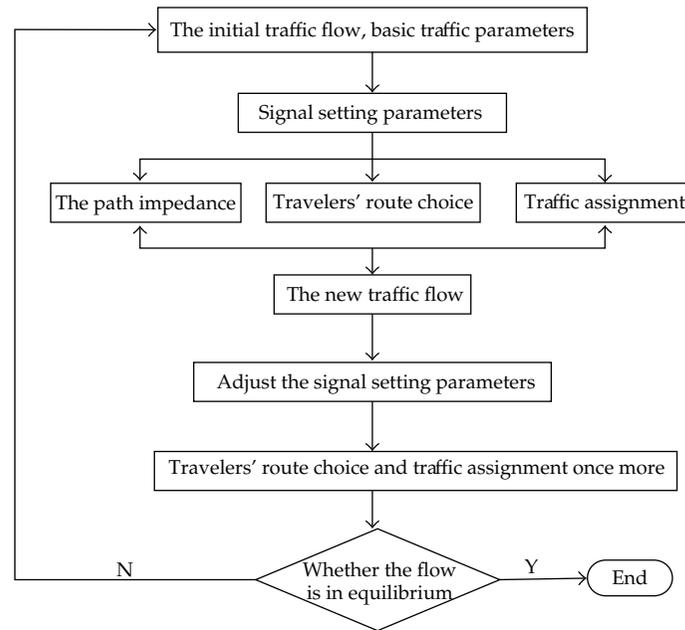


Figure 1: The relationship between signal setting and flow distribution.

The process of algorithm analysis is showed in Figure 1, which is summarized as follows.

Step 1. Obtain the initial traffic flow information based on real-time detection.

Step 2. Determine a set of signal setting parameters based on the initial traffic flow.

Step 3. Adjust the impedance of each path in the road network by traffic control, thereby affecting travelers' route choice.

Step 4. Calculate the flow after superimposed according to the travelers' route choice and the initial traffic flow.

Step 5. Adjust the signal setting parameters and traffic assignment again.

Step 6. Through the system utility value, determine the equilibrium situation.

The second part of the algorithm is to decide the signal setting parameters in order to balance the traffic flow under the influence of travelers' route rechoice. The main idea of this part is to determine the maximum green (g_{\max}) time or the minimum green time (g_{\min}) of each signal phase and then to extend or shorten the signal cycle (C) and green time (g) to ensure that the flow ratio (y_i) and the split (λ_i) are approximately equal. The signal control method under the impact of changing flow is shown in Figure 2.

The process of algorithm analysis is shown in Figure 2, which is summarized as follows.

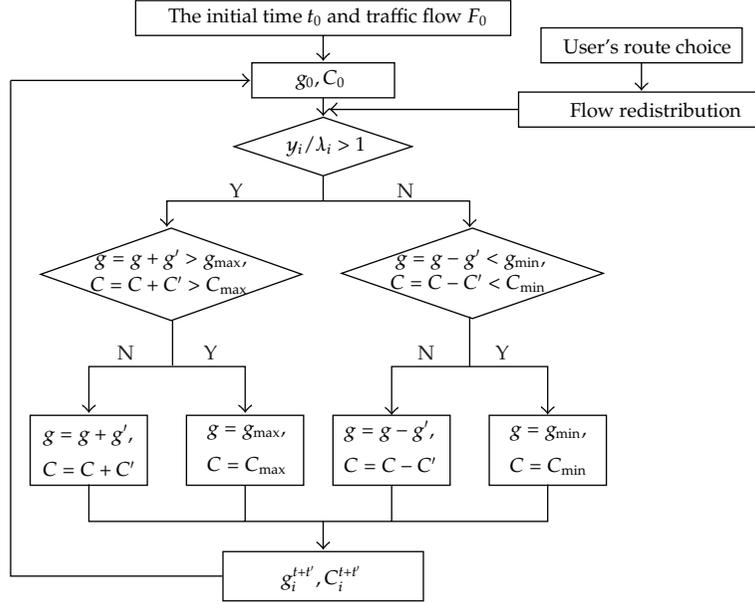


Figure 2: Signal control method under the impact of changing flow.

Step 1. Obtain the initial traffic flow information based on real-time detection.

Step 2. Calculate the initial signal cycle C_0 and the initial green time g_0 according to the following equation:

$$C_0 = \frac{1.5t_L + 5}{1 - Y}, \quad (3.1)$$

$$g_0 = (C_0 - t_L) \cdot \frac{y_i}{Y},$$

where t_L is the total lost time of intersection; Y is the flow ratio of intersection.

Step 3. Combined with the redistribution flow of road network to judge whether y_i/λ_i is approximate to 1, we can calculate the y_i/λ_i according to the following equation [20]:

$$\frac{y_i}{\lambda_i} = \frac{V_i(t_{ri} + g_i)}{V_{Si}g_i}, \quad (3.2)$$

where V_i is the average flow of phase i in critical lane; t_{ri} is the time interval of phase i from green time end to green time start; V_{Si} is the saturation flow of phase i in critical lane.

Step 4. Determine the C_{\max} , C_{\min} , g_{\max} , and g_{\min} for Step 5 to determine the suitable C and g . Generally, the cycle and the green time should not increase too much or too little; for this part we determine the C_{\max} and C_{\min} based on the following equation:

$$C_{\max} = \frac{4C_0}{3} = \frac{6t_L + 20}{3(1 - Y)}, \quad (3.3)$$

$$C_{\min} = \frac{0.9t_L}{(0.9 - Y)}.$$

Then we determine the g_{\max} , g_{\min} under the constraints of C_{\max} , C_{\min} ; the specific calculation method can be expressed as follows:

$$g_{i \max} = g_{i0} + g_{y \max} = C_{\max} - T_L - \sum_{j=1}^n g_{ir}, \quad (9)$$

where g_{ir} is the real green time of phase i ; $g_{y \max}$ is the maximum green time of phase i which can be increased:

$$g_{i \min} = g_{ir} - g_{ic} = \frac{(C_0 - L)y_{ir}}{Y}, \quad (3.4)$$

where g_{ic} is the compression time of phase i ; y_{ir} is the real green time of phase i .

Step 5. Increase or reduce the signal cycle and green time according to traffic demand under the constraints of C_{\max} , C_{\min} .

Step 6. Feed back the determined signal cycle and green time to Step 2.

4. Algorithm Implementation

4.1. Supply and Demand Balance Analysis

The supply of the road network mainly refers to the maximum capacity that can be provide under fixed road conditions, and the demand of road network refers to the number of users in traffic network. The traffic supply and demand can be expressed as follows:

$$\begin{aligned} Q_s &= \min \sum_{i \in p, l \in L} \delta_k^{rs} \lambda_i C_i, \\ Q_d &= Q_0 + \sum_{l \in L} \int_0^t [r_{ul}(t) - r_{vl}(t)] dt, \\ \text{s.t. } &Q_0 \geq 0, r_{uk}(t) \geq 0, r_{vk}(t) \geq 0, t \in [0, T], 0 \leq \delta_k^{rs} \leq 1, \end{aligned} \quad (4.1)$$

where Q_s , Q_d are traffic supply and traffic demand; δ_k^{rs} is the intersection correlation coefficient of path k from r to s ; Q_0 is the initial traffic flow of road network; $r_{uk}(t)$, $r_{vk}(t)$ are inflow and outflow rate of road network:

$$dd(t) = \frac{Q_d(t)}{Q_s(t)}, \quad (4.2)$$

where $dd(t)$ is equalization index.

When the relationship between supply and demand meets $dd(t) < 0.95$, traffic managers should shorten part of the signal cycle length in order to reduce the delay time. When the relationship between supply and demand meets $0.95 \leq dd(t) < 1.05$, we think that it is equilibrium of supply and demand. When the relationship between supply and demand meets $dd(t) \geq 1.05$, traffic managers should adjust the signal setting parameters in order to change the relationship between supply and demand [21].

4.2. User's Route Choice Model

For most users prefer the small impedance path to travel, based on this, we use user's route choice model. The conventional path impedance mainly includes link travel time and intersection delay. This paper uses an empirical formula considering the mixed traffic running characteristic based on BPR road impedance function to calculate link travel time and uses a simplified formula, Webster's delay function, to calculate intersection delay. The specific calculation model can be expressed as follows:

$$t_l = t_0 \left[1 + \alpha_1 \left(\frac{V_1}{C_1} \right)^{\beta_1} \right] \left[1 + \alpha_2 \left(\frac{V_2}{C_2} \right)^{\beta_2} \right], \quad (4.3)$$

$$t_j = \frac{c(1-\lambda)^2}{2(1-\lambda X)},$$

where t_l is travel time between adjacent intersections; t_j is the delay of intersection; t_0 is travel time under free speed between adjacent intersection; V_1, V_2 are the volumes of vehicle and nonmotorized vehicle; C_1, C_2 are the capacities of vehicle and nonmotorized vehicle; c, λ are the signal cycle and split of the intersection; X is intersection saturation.

Combined with (3.3), we can get path impedance as follows:

$$t_{ki}^{r-s} = t_0 \left[1 + \alpha_1 \left(\frac{V_1}{C_1} \right)^{\beta_1} \right] \left[1 + \alpha_2 \left(\frac{V_2}{C_2} \right)^{\beta_2} \right] + \frac{c(1-\lambda)^2}{2(1-\lambda X)}. \quad (4.4)$$

After determining the calculation method of path impedance, we need to research the relationship between the path impedance and the user's route choice. Dial (1971) proposed a route choice model which assumed that path impedance and route choice are independent of each other and obey the Gumbel probability distribution, but other path impedances are usually affected when one of path impedances changes. Thus, Wang Wei did some research and proved the model. This paper calculates the user's route choice based on the proved model, which is expressed as follows:

$$p_k^{rs} = \frac{e^{-\theta \cdot t_k^{rs} / \bar{t}_{rs}}}{\sum_{i=1}^n e^{-\theta \cdot t_i^{rs} / \bar{t}_{rs}}} \quad \forall k, r, s, \quad (4.5)$$

where p_k^{rs} is the probability of path k is chosen between origin (r) and destination (s); t_i^{rs} is the impedance of path k from origin (r) to destination (s); \bar{t}_{rs} is the average impedance of path k from origin (r) to destination (s).

4.3. Utility Functions Establishment

We can find that speed decreases with increasing flow in unsaturated and flow and speed decrease in saturated from speed-volume graph (Figure 3(a)); meanwhile, we also can find the maximum product of flow and speed from speed-volume graph (Figure 3(b)).

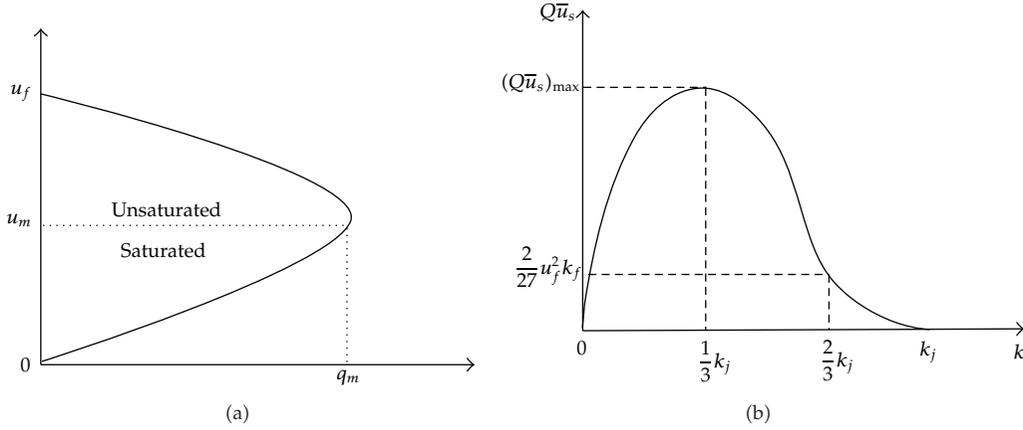


Figure 3: The speed-volume graph.

By analyzing Figure 3, for the realization of the maximum number of vehicles passing per unit of time, we consider the product of the flow and speed as the system utility value function, which can be expressed as follows:

$$\begin{aligned}
 t_k^{rs} &= t_{k1}^{rs} + t_{k2}^{rs} + t_{k3}^{rs} + \cdots + t_{kn}^{rs} = \sum_{m=1}^n t_{km}^{rs} \\
 q_k^{rs} &= q_u P_k^{rs} = q_u \frac{1}{1 + \sum_{i=1, i \neq k} e^{-\theta \cdot (t_i^{rs} / \bar{t}_{rs} - t_k^{rs} / \bar{t}_{rs})}}, \\
 q_k^{rs} &= \int_0^t [r_{uk}(t) - r_{vk}(t)] dt, \\
 \max J^{rs} &= \max \sum (q_{k0}^{rs} + q_k^{rs}) u_k^{rs} = \sum q_{km}^* \frac{l_k}{t_k^{rs*}},
 \end{aligned} \tag{4.6}$$

where J^{rs} is the system utility value from origin (r) to destination (s); q_u is the inflow volume of road network; q_k^{rs} is the difference between inflow and outflow; q_{k0}^{rs} is the initial traffic flow of path k ; l_k is the length of path k from origin (r) to destination (s); q_{km}^* is the optimal flow of path k ; t_k^{rs*} is the optimal impedance of path k .

4.4. User's Equilibrium Traffic Assignment Discriminant Model

Since we have already known the path impedance and the user's route choice and have established a utility function, we can summarize the above analysis results to discriminate the user's equilibrium degree of road network based on the following formulas:

$$\begin{aligned}
 dd(k) &= \frac{t_k^{rs*}}{t_k^{rs}}, \\
 dd(v) &= \frac{dd(k)}{dd(k+i)}, \\
 dd(v) &= \begin{cases} 1, & \text{if } |dd(v)| \leq \tau, \\ 0, & \text{otherwise,} \end{cases}
 \end{aligned} \tag{4.7}$$

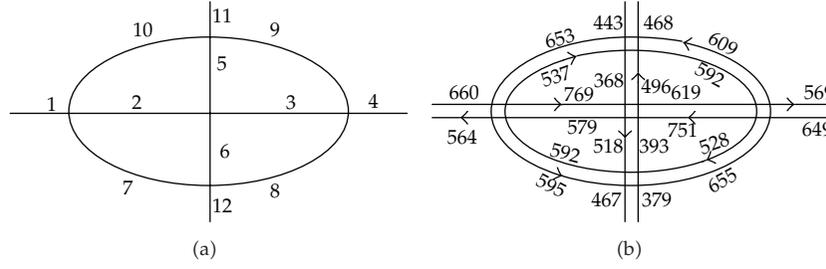


Figure 4: The number and the traffic flow of links.

Table 1: The links length.

Number	1	2	3	4	5	6	7	8	9	10	11	12
Length	532	1118	1244	478	593	616	1494	1566	1609	1428	336	314

where $dd(k)$ is the equilibrium degree of path k ; $dd(v)$ is the equilibrium degree between path k and $(k + i)$; τ is a fixed value where $0 \leq \tau \leq 1$.

5. Numerical Example

The test network is a simplified region of road network from Changchun city which consists of 12 links and 5 intersections. All links of the road network are bidirectional and of four lanes, and the initial data is obtained by manual collection. The number of links in the road network is shown in Figure 4(a), and the road link' traffic flow is shown in Figure 4(b). The links length is shown in Table 1.

We can find four pairs of O-D which, respectively, belong to link 1, link 4, link 11, and link 12. This paper simulated a pair of O-D (from link 1 to link 4) to determine the equilibrium degree of road network and to verify the superiority of the coordination algorithm. There are seven paths from link 1 to link 4, that is, path A (1-2-3-4), path B (1-7-8-4), path C (1-7-6-3-4), path D (1-10-9-4), path E (1-10-5-3-4), path F (1-2-5-9-4), and path G (1-2-6-8-4). We can find that the shortest distance is the path 1-2-3-4, but there is a heavy traffic flow on link 2 and link 3.

In the simulation test process, we set the traffic signal based on Synchro Studio 7 which is a traffic signal design software; we also have embedded the above algorithms into the traffic signal design software. During the inflow increase from 1600 pcu/h to 4400 pcu/h, we first get the path impedance based on VISSIM traffic simulation software which is shown in Table 2.

Then we can get the optimal impedance, the optimal flow, and the optimal utility value based on formula (4.2) as follows (the specific data is shown in Table 3 and the relationship between utility value and flow is shown in Figure 5):

$$J^{1-4} = \max \sum q_{km}^* \frac{l_k}{t_k^{1-4*}} = 116485. \tag{5.1}$$

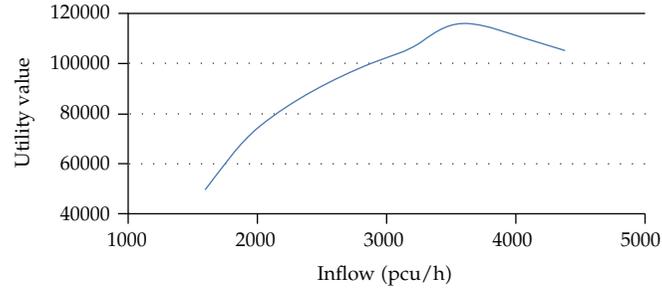


Figure 5: The relationship between utility value and inflow.

Table 2: The path impedance based on simulation data.

Volume (pcu/h)	Number						
	Path impedance after application of algorithm(s)						
	A	B	C	D	E	F	G
1600	315	311	340	334	387	352	368
2000	324	333	368	359	422	378	401
2400	356	359	391	385	451	405	425
2800	395	402	436	425	489	443	461
3200	431	426	473	447	536	479	508
3600	488	458	525	486	593	562	547
4000	609	551	646	588	704	625	652
4400	637	654	716	654	729	724	718

Table 3: The optimal flow and speed.

	A	B	C	D	E	F	G
q_{km}^* (pcu/h)	581	538	493	526	507	497	486
u_k^{1-4*} (km/h)	26	31	35	29	37	35	33
t_k^{1-4*} (s)	467	472	449	503	416	446	470

We next verify the superiority of the algorithm by inputting the initial data into the test network. During the simulation, we should mark all vehicles in order to record the different paths' impedance and then obtain the average impedance of each path. The average impedance and the equilibrium degree are shown as follows.

As seen in Tables 4 and 5, the new algorithm achieved the equilibrium of road network and that could save 21 seconds for each vehicle from point 1 to point 4 in the test network. Although this test network demonstrates a better performance of the coordination algorithm, more extensive tests need to be performed to validate the algorithm, for the above test only simulated a pair of O-D and did not consider the impact of distance on the users' route choice.

Table 4: Before application of algorithms.

	A	B	C	D	E	F	G
Paths impedance	494	423	518	466	502	487	496
Equilibrium degree(s)	0.945126	1.117364	0.866541	1.07808	0.828577	0.914520	0.947947
Average impedance(s)				484			

Table 5: After application of algorithms.

	A	B	C	D	E	F	G
Paths impedance	483	474	438	506	433	454	456
Equilibrium degree(s)	0.966650	0.997141	1.024814	0.992858	0.960614	0.980994	1.031100
Average impedance(s)				463			

6. Conclusions

In this paper, we presented a traffic control algorithm considering traffic assignment of road network, which has been implemented successfully by introducing the link impedance function and the route choice function. In the process of optimizing the coordination algorithms, we also get the optimal flow of path, the optimal impedance, the maximum system utility value, and so forth. Finally, we verified the superiority of the coordination algorithm and the model through a numerical example based on the VISSIM software.

For further issues, we should introduce traffic guidance information into the coordination algorithm of traffic control and traffic assignment and then verify it by the road network with multiorigin and multideestination to be closer to the actual situation.

Acknowledgments

This work is partly supported by the National Science Foundation of China (Grant nos. 50908100, 51108208), China Postdoctoral Science Foundation funded Project (no. 20110491307), and the Fundamental Research Funds for the Central Universities (no. 201103146). The authors thank the anonymous reviewers for their valuable input and suggestions.

References

- [1] H. Qi, *Research on the Modelling of Network Traffic Flow Considering Signal Control*, Jilin University, Jilin, China, 2011.
- [2] S. Hu, *Study on the Equilibrium Theory between Traffic Supply and Traffic Demand*, Changan University, Shanxi, China, 2004.
- [3] R. E. Allsop, "Some possibilities for using traffic control to influence trip distribution and route choice," in *Transportation and Traffic Theory*, pp. 345–373, Elsevier, New York, NY, USA, 1974.
- [4] N. H. Gartner, "Area traffic control and network equilibrium," in *Proceedings of the International Symposium on Traffic Equilibrium Methods*, pp. 274–297, Berlin, Germany, 1976.
- [5] H. N. Tan, S. B. Gershwin, and M. Athans, "Hybrid optimization in urban traffic networks," LIDS Technical Report, MIT Press, Cambridge, Mass, USA, 1979.
- [6] M. J. Smith, "The existence, uniqueness and stability of traffic equilibria," *Transportation Research Part B*, vol. 13, no. 4, pp. 295–304, 1979.
- [7] M. J. Smith, "A local traffic control policy which automatically maximises the overall travel capacity of an urban road network," *Traffic Engineering and Control*, vol. 21, no. 6, pp. 298–302, 1980.

- [8] N. H. Gartner, S. B. Gershwin, J. D. C. Little, and P. Ross, "Pilot study of computer-based urban traffic management," *Transportation Research B*, vol. 14, no. 1-2, pp. 203–217, 1980.
- [9] C. S. Fisk, "Game theory and transportation systems modelling," *Transportation Research Part B*, vol. 18, no. 4-5, pp. 301–313, 1984.
- [10] H. Yang and S. Yagar, "Traffic assignment and signal control in saturated road networks," *Transportation Research A*, vol. 29, no. 2, pp. 125–139, 1995.
- [11] N. H. Gartner and C. Stamatiadis, "Framework for the integration of dynamic traffic assignment with real-time control," in *Proceedings of the 3rd Annual World Congress on Intelligent Transportation Systems*, pp. 125–137, Orlando, Fla, USA, 1996.
- [12] N. H. Gartner and M. Al-Malik, "Combined model for signal control and route choice in urban traffic networks," *Transportation Research Record*, vol. 1554, pp. 27–35, 1996.
- [13] C. Lee and R. B. Machemehl, "Genetic algorithm, local and iterative searches for combining traffic assignment and signal control," in *Proceedings of the Conference on Traffic and Transportation Studies (ICTTS '98)*, pp. 27–29, Beijing, China, July 1998.
- [14] O. J. Chen and M. E. Ben-Akiva, "Game-theoretic formulations of interaction between dynamic traffic control and dynamic traffic assignment," *Transportation Research Record*, no. 1617, pp. 179–188, 1998.
- [15] S. W. Chiou, "TRANSYT derivatives for area traffic control optimisation with network equilibrium flows," *Transportation Research B*, vol. 37, no. 3, pp. 263–290, 2003.
- [16] E. Cipriani and G. Fusco, "Combined signal setting design and traffic assignment problem," *European Journal of Operational Research*, vol. 155, no. 3, pp. 569–583, 2004.
- [17] E. Cascetta, M. Gallo, and B. Montella, "Models and algorithms for the optimization of signal settings on urban networks with stochastic assignment models," *Annals of Operations Research*, vol. 144, pp. 301–328, 2006.
- [18] H. Gao, T. Chen, and J. Lam, "A new delay system approach to network-based control," *Automatica*, vol. 44, no. 1, pp. 39–52, 2008.
- [19] S.-W. Chiou, "Optimization of a nonlinear area traffic control system with elastic demand," *Automatica*, vol. 46, no. 10, pp. 1626–1635, 2010.
- [20] Y. Bie, D. Wang, X. Song, and Y. Xing, "Conditional bus signal priority strategies considering saturation degree restriction at isolated junction," *Journal of Southwest Jiaotong University*, vol. 46, no. 4, pp. 657–663, 2011.
- [21] L. Binglin and K. Aiwu, "Nonequilibrium study on the supply and demand of urban road networks," *Shanxi Science and Technology*, pp. 58–60, 2008.

Research Article

Estimation of Saturation Flow Rates at Signalized Intersections

Chang-qiao Shao and Xiao-ming Liu

Key Lab of Transportation Engineering, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Chang-qiao Shao, shaocq@bjut.edu.cn

Received 26 August 2012; Revised 11 October 2012; Accepted 27 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 C.-q. Shao and X.-m. Liu. 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 saturation flow rate is a fundamental parameter to measure the intersection capacity and time the traffic signals. However, it is revealed that traditional methods which are mainly developed using the average value of observed queue discharge headways to estimate the saturation headway might lead to underestimate saturation flow rate. The goal of this paper is to study the stochastic nature of queue discharge headways and to develop a more accurate estimate method for saturation headway and saturation flow rate. Based on the surveyed data, the characteristics of queue discharge headways and the estimation method of saturated flow rate are studied. It is found that the average value of queue discharge headways is greater than the median value and that the skewness of the headways is positive. Normal distribution tests were conducted before and after a log transformation of the headways. The goodness-of-fit test showed that for some surveyed sites, the queue discharge headways can be fitted by the normal distribution and for other surveyed sites, the headways can be fitted by lognormal distribution. According to the queue discharge headway characteristics, the median value of queue discharge headways is suggested to estimate the saturation headway and a new method of estimation saturation flow rates is developed.

1. Introduction

The saturation flow rate is a fundamental parameter to measure the intersection capacity and time the traffic signals. Traditional saturation flow rate estimation methods were mostly developed based on the assumption that the queue discharge headway is a fairly constant and that the average headway estimated from the first 4-to-10 or 4-to-12 vehicles is representative of the saturation headway (h) meaning every vehicle in a stable moving platoon consumes h seconds [1, 2]. However, variability in queue discharge headways is addressed in more recent studies [2–14] and it is found that traditional methods which simply use the average of discharge headway to estimate the saturation headway might lead to underestimate

saturation flow rate. Errors in saturation flow rates used to estimate vehicle delays carry over onto delay predictions and level of service (LOS) predictions. Therefore, it is necessary to study and to improve accuracy of the estimation of saturation flow rate. Due to the fact that driver behavior is the main determinate of saturation flow rate, a good understanding of the discharge headway of individual vehicles is essential in estimation of saturation headway and saturation flow rate.

However, the distribution characteristic of headways and its effect on the estimation of saturation flow rate have not been examined in literatures. The goal of this paper is to investigate the stochastic nature of queue discharge headways, which may provide better information, and, so, to develop a new saturation flow rate estimation method. To fulfill the objective, a literature review, data collection and data analysis, and model development were conducted.

2. Literature Review

Due to the importance of the queue discharge headways and the saturation flow rates in intersection capacity estimation and traffic control, researches have made great efforts to study the characteristics of discharge headways and saturation flow rates. Li and Prevedouros [2] applied three methodologies to analyze the saturation headway based on the field data. It was revealed that when long queues are present, the typical field measurement of saturation flow rate based on the first 12 vehicles is an overestimate of saturation flow rate for through vehicles and an underestimate of saturation flow rate for protected left-turning vehicles. Lu and Pernía [3] analyzed the differences of driving behavior among different driver age groups based on collected data at signalized intersections. And the variation in capacities with an increasing percentage of older drivers in the traffic stream was modeled. It was found that the presence of older drivers significantly reduced intersection capacity at all study sites because of their higher lost times and lower saturation flow rates. Hung et al. [4] performed one study on the discharge headways at 26 sites of signalized intersections in Hong Kong. It is concluded that discharge headway at different queue position follows the type I extreme value distribution. Mei and Bullen [5] analyzed the individual headways distribution for drivers in a car-following state and it was indicated that the lognormal distribution yielded a good fit for headways in busy traffic. Lin and Thomas [7] examined the queue discharge characteristics at three intersections on Long Island, New York. It is found that queue discharge rates often kept rising even after the 15th vehicle has entered the intersection. Jin et al. [8] also found that the distributions of the departure headways at each position in a queue approximately follow a certain log-normal distribution. Joseph and Chang [10] illustrated the field measurements of saturation flow headways and start-up lost times and found that they varies significantly ranging from 1.63 to 1.91 s. Long [11] indicated that traditional estimates of saturation flow rates have been challenged by field observations, and that large variations in saturation flow measurements have cast doubt on the presumption that the base saturation flow rate is either stable or a constant value. Rahman et al. [12] developed a new procedure for determining the saturation flow rates at signalized intersections and they concluded that the procedure as discussed in the HCM [1] would overestimate saturation flow rates for some approaches and underestimate them for the other approaches based on the field data. Shao et al. [13] reported some findings on the saturation flow rates at signalized intersections in China. It is shown that the average headway is about 2.03 s and the standard error is from 0.40 s to 0.90 s. Sharma et al. [14] studied queue

discharge characteristics at signalized intersections under heterogeneous traffic conditions and the effect of a countdown timer on the headway distribution. The results indicated that the accepted headway distribution is followed when there is no timer.

From the efforts of these researches, it can be concluded that the queue discharge headway is not constant due to various factors and that the effect of headway variations on the estimation of saturation flow rate should be considered.

3. Queue Discharge Headway and Saturation Flow Rate

Discharge headways at signalized intersections are defined as the time intervals between two successive vehicles passing a stop line or any predetermined reference line at the intersections. It can be expressed as the function of the driver response during queue discharge [15], spacing between vehicle in queue and queue discharge speed:

$$h_s = t_r + \frac{L_{hj}}{v_s}, \quad (3.1)$$

where h_s is the discharge headway(s); t_r is the driver response time (s); L_{hj} is queue space per vehicle (include vehicle length, m); v_s is the queued vehicle speed (m/s). From (3.1), it is easy to understand that the queue discharge headway h_s is a random variable due to the difference of driver behavior (response time t_r and variation in the length of queue space L_{hj} [2, 15]). In order to estimate the saturation headway, the average queue discharge headway is calculated as follows:

$$\bar{h}_s = \frac{1}{n} \sum_{i=1}^n h_{si}, \quad (3.2)$$

where, \bar{h}_s is the average queue discharge headway(s) which can be taken as the estimation of saturation headway h ; h_{si} is the headway of vehicle i in the queue, $i = 1, 2, \dots, n$ (n is the sample size). And, consequently, the traditional saturation flow rate is determined by the queue discharge headway:

$$S = \frac{3600}{\bar{h}_s}, \quad (3.3)$$

where, S is the saturation flow rate (veh/h). According to the knowledge of statistics [16], when the distribution of queue discharge headways is symmetrical, the average headway calculated by (3.2) is unbiased estimation of h and (3.3) can be a good estimation of the saturation flow rate. However, when the headway distribution is unsymmetrical, \bar{h}_s is not a good estimation and (3.3) will overestimate or underestimate saturation flow rate. Therefore, it is necessary to learn about the stochastic nature of queue discharge headway and to analyze the effect of variation of headways on the estimation of saturation flow rate.

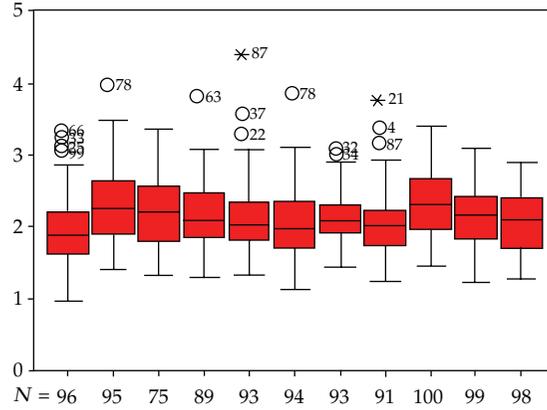


Figure 1: Boxplot of queue discharge headways.

4. Methodology

4.1. Data Collection

Discharge headway data were collected at 11 signalized intersections in Beijing. Surveys were conducted at the morning and evening peak periods (7:00 to 9:00 and 17:00 to 18:00) at weekdays by video cameras and the data were manually collected from the videotapes. Through the videotaping, the time for each vehicle rear bumper passing the stop line was recorded and the time headway was calculated.

In order to eliminate the effect of start-up and acceleration on the saturation flow rate, the first 5 headways in each signal cycle were removed from the data. Also, data which were apparently anomalous because the drivers are in distracted attention that causes the headway exceeded the normal ranges are eliminated. In addition, the signal cycle in which vehicles were seriously disturbed by pedestrians or bicycles is not included. A total of 1023 vehicle discharge headways were recorded.

4.2. Data Analysis

In order to investigate the queue discharge headway statistical characteristics, the surveyed headways at 11 sites of signalized intersections are analyzed (Table 1 presented the results). It is easily found that there are some differences in the queue discharge headways for different surveyed sites (Figure 1). However, there are still some interesting characteristics which should be noted. The average value of queue discharge headways is greater than the median value (the difference range from 0.0 s to 0.30 s) which is consistent with the existing findings [6, 12, 13]. Also, it is found that the skewness of headways is positive. These characteristics indicate that the distribution of queue discharge headway is likely unsymmetrical and the normal distribution function is not appropriate to fit the headway data.

The fact that average value of queue discharge headway \bar{h}_s is greater than the median value shows that more than 50% drivers will keep smaller headway than \bar{h}_s . Therefore, traditional saturation flow rate estimation method would underestimate the flow rates.

Table 1: Statistical results for the queue discharge headways.

Code of surveyed site	Sample size	Average	Median	Stand error	Skewness	Kurtosis
1	96	1.95	1.88	0.52	0.66	0.88
2	95	2.36	2.28	0.55	0.61	1.22
3	75	2.30	2.23	0.49	0.43	-0.10
4	89	2.22	2.14	0.37	0.88	1.09
5	93	2.19	2.03	0.43	0.97	1.21
6	94	2.08	1.98	0.47	0.43	-0.34
7	93	2.21	2.09	0.44	0.35	-0.39
8	91	2.04	2.00	0.43	0.81	1.11
9	100	2.31	2.30	0.45	0.02	-0.75
10	99	2.18	2.16	0.46	0.07	-0.73
11	98	2.11	2.07	0.42	0.02	-0.20

5. Discharge Headway Distribution

According to the statistical analysis result of the surveyed queue discharge headway data, it is revealed that the sample average value is greater than median value and that queue discharge headway distribution is likely unsymmetrical. Recent research [5, 8] showed that the queue discharge headway could be fitted by lognormal distribution. So, in this paper, the lognormal distribution is chosen to fit the queue discharge headway data (as comparison, the normal tests for time headway data are also conducted). The lognormal distribution density function is

$$f_H(h) = \frac{1}{\sqrt{2\pi}\sigma h} \exp\left(-\frac{(\ln h - u)^2}{2\sigma^2}\right), \quad h \geq 0. \quad (5.1)$$

The means, median, and variance of lognormal distribution are the following:

$$E(H) = \exp(\mu + 0.5\sigma^2), \quad (5.2)$$

$$\text{Med}(H) = \exp(u), \quad (5.3)$$

$$\text{Var}(H) = \exp(2\mu + \sigma^2)(e^{\sigma^2} - 1). \quad (5.4)$$

It is known that when the random variable h_s density function is lognormal, then $\ln(h_s)$ will follow the normal distribution. Applying this quality of lognormal distribution, the hypothesis that discharge headway h_s is lognormal distributed can be tested using normal test method (in this study, the Shapiro-Wilk method is used [17]). Table 2 listed the hypothesis test results. From Table 2, it is concluded that among the 11 surveyed sites, there are 7 signalized intersections at which queue discharge headways do not pass the normal distribution hypothesis test (with significant level $P(W \leq w) \leq 0.05$) while they pass the lognormal distribution hypothesis test (with significant level $P(W^* \leq w^*) > 0.05$). On the contrary, there are 3 signalized intersections at which queue discharge headways pass the normal distribution hypothesis test ($P(W \leq w) > 0.05$) but they do not pass the lognormal

Table 2: Result of headway distribution test.

Code of surveyed site	1	2	3	4	5	6	7	8	9	10	11
w	0.952	0.953	0.968	0.952	0.941	0.955	0.969	0.947	0.968	0.965	0.974
$P(W \leq w)$	0.005	0.007	0.100	0.006	0.0006	0.012	0.069	0.003	0.116	0.064	0.239
w^*	0.981	0.972	0.976	0.982	0.979	0.971	0.978	0.976	0.911	0.955	0.960
$P(W^* \leq w^*)$	0.601	0.209	0.448	0.655	0.499	0.177	0.147	0.342	0.025	0.007	0.021

w and w^* are Shapiro-Wilk statistics, w^* is computed after log transformation.

distribution hypothesis test ($P(W^* \leq w^*) < 0.05$). There is only one site (code number is 3) which significant level of headway distribution being normal is 0.10 and significant level of being lognormal is 0.448.

Therefore, among the surveyed 11 signalized intersections, there are 8 signalized intersections at which headway distribution is lognormal while the other 3 signalized intersections at which headway distribution is normal.

6. New Estimation of Saturation Flow Rate

6.1. Estimation of Saturation Flow Rate Based on the Median Value of Queue Discharge Headways

According to the analysis results of surveyed queue discharge headway data, the fact that for some signalized intersections, the queue discharge headways can be fitted by lognormal distribution means that the simple average of headway is not a good estimation and (3.3) will underestimate the saturation flow rate. Hence, it is reasonable to use the median value [16] to calculate the saturation flow rate and a new estimation is obtained:

$$S_1 = 3600 \times \frac{1}{h_{\text{med}}}, \quad (6.1)$$

where S_1 is a new estimation of saturation flow rate (veh/h); h_{med} is sample median value of headways or estimation of median headways. For the lognormal distribution (5.1), the likelihood estimations for parameters μ and σ^2 are noted as $\hat{\mu}_{\text{ML}}$ and $\hat{\sigma}_{\text{ML}}^2$, respectively. According to the statistical knowledge [16], $\hat{\mu}_{\text{ML}}$ and $\hat{\sigma}_{\text{ML}}^2$ can be expressed as

$$\hat{\mu}_{\text{ML}} = \frac{1}{n} \sum_{i=1}^n \ln h_{si}, \quad (6.2)$$

$$\hat{\sigma}_{\text{ML}}^2 = \frac{1}{n} \sum_{i=1}^n \left[\ln h_{si} - \frac{1}{n} \sum_{i=1}^n \ln h_{si} \right]^2. \quad (6.3)$$

Let $\hat{\mu}_M$ and $\hat{\sigma}_M^2$ denote the moment estimations for parameters μ and σ^2 . It is easily derived that

$$\hat{\mu}_M = \ln \left(\frac{\bar{h}_s}{\sqrt{1 + \sigma_s^2 / \bar{h}_s^2}} \right), \quad (6.4)$$

$$\hat{\sigma}_M^2 = \ln \left(1 + \frac{\sigma_s^2}{\bar{h}_s^2} \right). \quad (6.5)$$

From (5.3), (6.2), and (6.4), other two new estimations of saturation headway are obtained:

$$\begin{aligned} \widehat{\text{Med}}(h_s)_1 &= \exp(\hat{u}_{\text{ML}}) = \exp \left(\frac{1}{n} \sum_{i=1}^n \ln h_{si} \right), \\ \widehat{\text{Med}}(h_s)_2 &= \exp(\hat{u}_M) = \frac{\bar{h}_s}{\sqrt{1 + \sigma_s^2 / \bar{h}_s^2}}. \end{aligned} \quad (6.6)$$

Consequently, two new estimations of saturation flow rate are also suggested:

$$S_2 = 3600 \times \exp \left(-\frac{1}{n} \sum_{i=1}^n \ln h_{si} \right), \quad (6.7)$$

$$S_3 = \frac{3600}{\bar{h}_s} \times \left(1 + \frac{\sigma_s^2}{\bar{h}_s^2} \right)^{0.5}, \quad (6.8)$$

where S_2 and S_3 are two new estimations of saturation flow rate (veh/h); $\sigma_s^2 = [(1/n - 1) \sum_{i=1}^n (h_{si} - \bar{h}_s)^2]$ is the variance of sample headways.

Equation (6.7) is developed based on the likelihood estimation of median value and it is consistent with (6.1). Equation (6.8) includes information from both the average value of queue discharge headway \bar{h}_s and variance σ_s^2 . (Hence, the estimation takes into account the effect of queue discharge headway spread on the estimation.) It can be taken as the modification of (3.3).

6.2. Comparison of Saturation Flow Rate Estimations

Table 3 listed the saturation flow rates based on different estimation methods by using the surveyed data at 11 signalized intersections. For the first 8 signalized intersections (their queue discharge headways can be taken as being from lognormal distribution), the saturation flow rates are computed, respectively, using these methods developed in this paper. It can be found that the computed saturation flow rates based on median of queue discharge headways (6.1), (6.7), and (6.8) are very close to each other and higher than the values estimated by the traditional method (3.3) for each surveyed sites. For the other 3 signalized intersections

Table 3: Comparison of computed saturations using different methods.

Code of surveyed site	Headway means (s)	Headway median (s)		Saturation flow rates (veh/h)			
		Sample median	Equation (6.2)	Equation (3.3)	Equation (6.1)	Equation (6.7)	Equation (6.8)
1	1.95	1.88	1.88	1846	1914	1914	1911
2	2.36	2.28	2.25	1525	1578	1600	1656
3	2.30	2.23	2.18	1565	1614	1651	1600
4	2.22	2.14	2.11	1621	1682	1706	1644
5	2.19	2.03	2.06	1643	1773	1747	1675
6	2.08	1.98	1.99	1730	1818	1809	1774
7	2.21	2.09	2.12	1558	1723	1698	1660
8	2.04	2.00	2.00	1764	1800	1800	1803
9	2.31	2.30	2.27	1558	1565	—	—
10	2.18	2.16	2.05	1651	1666	—	—
11	2.11	2.07	2.06	1706	1739	—	—

(their queue discharge headways can be taken as being from normal distribution), the saturation flow rates estimated by the method (6.1), developed in this paper, is near the value obtained by using the traditional method. Hence, these findings validated that when the headway distribution is unsymmetrical (e.g., lognormal distribution is unsymmetrical), the traditional method will underestimate the saturation flow rates. Otherwise, when the distribution is symmetrical (normal distribution is symmetrical), the saturation flow rate estimation method developed in this paper is consistent with the traditional method.

7. Conclusions

This paper reported results of the queue discharge headway distribution at 11 surveyed signalized intersections and the following conclusions can be made

- (1) The average value of queue discharge headways is greater than the median value, and the skew of headways is positive.
- (2) The goodness-of-fit test showed that for some surveyed sites, the headways follow the normal distribution and that for other surveyed sites, the queue discharge headways are lognormal distributed.
- (3) The traditional estimation of saturation headway does not accurately reflect the true value of h and would overestimate it. As a result, the traditional estimation of saturation flow rate would underestimate saturation flow rate.
- (4) The new estimations of saturation flow rate developed in this study are more reasonable and they are suggested to be used in traffic control and measurement of intersection capacity.

In addition, it is found that for different signalized intersections, the queue discharge headway distribution can be not identical. Does it relate to the conditions or traffic circumstance? This is a new problem and needs to be verified.

Acknowledgments

This work was supported by the National Key Basic Research Program of China under Grant no. 2012CB723303. The authors would like to thank the anonymous reviewers whose comments helped to significantly improve the overall quality of this paper.

References

- [1] Transportation Research Board, *Highway Capacity Manual 2000*, National Research Council, Washington, DC, USA, 2000.
- [2] H. Li and P. D. Prevedouros, "Detailed observations of saturation headways and start-up lost times," *Transportation Research Record*, no. 1802, pp. 44–53, 2002.
- [3] J. J. Lu and J. C. Pernía, "The difference of driving behavior among different driver age groups at signalized intersections," *IATSS Research*, vol. 24, no. 2, pp. 75–84, 2000.
- [4] W. T. Hung, F. Tian, and H. Y. Tong, "Discharge headway at signalized intersections in Hong Kong," *Journal of Advanced Transportation*, vol. 37, no. 1, pp. 105–117, 2003.
- [5] M. Mei and A. G. R. Bullen, "Lognormal distribution for high traffic flows," *Transportation Research Record*, no. 1398, pp. 125–128, 1993.
- [6] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research Part C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [7] F. B. Lin and D. R. Thomas, "Headway compression during queue discharge at signalized intersections," *Transportation Research Record*, no. 1920, pp. 81–85, 2005.
- [8] X. Jin, Y. Zhang, F. Wang et al., "Departure headways at signalized intersections: a log-normal distribution model approach," *Transportation Research Part C*, vol. 17, no. 3, pp. 318–327, 2009.
- [9] R. Akçelik, "An investigation of the performance of roundabout with metering signals," Paper for Presentation at the National Roundabout Conference, Transportation Research Board, Kansas City, Kan, USA, 2008.
- [10] J. Joseph and G. L. Chang, "Saturation flow rates and maximum critical lane volumes for planning applications in Maryland," *Journal of Transportation Engineering*, vol. 131, no. 12, pp. 946–952, 2005.
- [11] G. Long, "Driver behavior model of saturation flow," *Transportation Research Record*, no. 2027, pp. 65–72, 2007.
- [12] M. Rahman, S. A. Nur-ud-deen, and T. Hassan, "Comparison of saturation flow rate at signalized intersections in Yokohama and Dhaka," *Proceedings of the Eastern Asia Society for Transportation Studies*, vol. 5, pp. 959–966, 2005.
- [13] C. Q. Shao, J. Rong, and X. M. Liu, "Study on the saturation flow rate and its influence factors at signalized intersections in China," *Procedia—Social and Behavioral Sciences*, pp. 504–514, 2011.
- [14] A. Sharma, L. Vanajakshi, and N. Rao, "Effect of phase countdown timers on queue discharge characteristics under heterogeneous traffic conditions," *Transportation Research Record*, no. 2130, pp. 93–100, 2009.
- [15] R. Akçelik and M. Besley, "Queue discharge flow and speed models for signalized intersections," in *Proceedings of the 15th International Symposium on Transportation and Traffic Theory*, pp. 99–118, Pergamon, Elsevier Science, Oxford, UK, 2002.
- [16] F. J. Wall, *Statistical Data Analysis Handbook*, Mcgraw-Hill, 1986.
- [17] J. A. Rice, *Mathematical Statistics and Data Analysis*, Brooks/Cole, 2007.

Research Article

An Optimization to Schedule Train Operations with Phase-Regular Framework for Intercity Rail Lines

Huimin Niu and Minghui Zhang

School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

Correspondence should be addressed to Huimin Niu, hmniu@mail.lzjtu.cn

Received 7 September 2012; Accepted 15 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 H. Niu and M. Zhang. 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 most important operating problem for intercity rail lines, which are characterized with the train operations at rapid speed and high frequency, is to design a service-oriented schedule with the minimum cost. This paper proposes a phase-regular scheduling method which divides a day equally into several time blocks and applies a regular train-departing interval and the same train length for each period under the period-dependent demand conditions. A nonlinear mixed zero-one programming model, which could accurately calculate the passenger waiting time and the in-train crowded cost, is developed in this study. A hybrid genetic algorithm associated with the layered crossover and mutation operation is carefully designed to solve the proposed model. Finally, the effectiveness of the proposed model and algorithm is illustrated through the application to Hefei-Wuhan intercity rail line in China.

1. Introduction

Intercity rail lines, as a rapid transport mode connecting two cities, have been paid much attention by the governments all over the world. They have become one of the most important engines for boosting regional economic development and accelerating the urbanization process. In recent years, great importance has been attached to their construction in China. The associated rail networks connecting many important cities have been built or are under construction in several economically developed regions, such as Pearl River zone and Yangtze River zone, and there is a current trend for them to expand to the rest of the country.

An intercity rail line, either in matters of passenger demands or operation scheduling, is different from the general one. For intercity rail lines, the operations are characterized with rapid speed and high frequency, and the process of passengers arriving at the stations are time-dependent and stochastic. An even schedule with a constant headway between consecutive trains may result in long passenger waiting times during oversaturated periods,

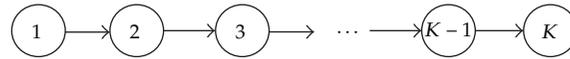


Figure 1: The illustration of an intercity rail line.

or ineffective train capacity utilization during unsaturated periods due to the undercapacity. A train schedule with inconstant headways, however, may lead to the frequent adjustment of predetermined schedule and the complexity of operations management. There is also trouble in determining the number of cars constituting a train, which should not be always steady or frequently modified.

Various attempts have been made to obtain a train schedule for railroad transports using optimization methods in the past decades. Ghoneim and Wirasinghe [1] defined the total cost as the sum of the travel time and the relevant rail capital and operating cost and designed a train scheduling plan with the minimum cost. Goossens et al. [2] introduced several models for solving operational scheduling problems in which railway lines can have different halting patterns. Liebchen [3] adapted a periodic event-scheduling approach and a well-established graph model to optimize the Berlin subway timetable. Claessens et al. [4] developed a mathematical programming model subjected to service and capacity constraints to optimize train operations. Ghoseiri et al. [5] built a multiobjective optimization model for the passenger train scheduling problem on a rail network which includes single and multiple tracks, as well as multiple platforms with different train capacities. Khan and Zhou [6] developed a stochastic optimization formulation for incorporating segment travel-time uncertainty and dispatching policies into a medium-term train-timetabling process, which is aimed to minimize the total trip time in a published timetable and reduce the expected schedule delay. Zhou and Zhong [7] formulated train scheduling models which consider segment and station headway capacities as limited resources, and developed algorithms to minimize both the expected passenger waiting times and total train travel times. Nachtigall and Voget [8] discussed the cost benefit between the investigation for reforming track states and the quality of the resulting timetable measured by the remaining waiting times. Palma and Lindsey [9] analyzed the schedule delay costs incurred from travelling earlier or later than desired and formulated an optimization model with the objective of minimizing the total riders' schedule delay costs. Nguyen et al. [10] presented a graph theoretic framework for the passenger assignment problem. Wong et al. [11] presented a mixed-integer-programming model for the schedule synchronization to minimize passengers' transfer times. Meng and Zhou [12] established a robust single-track train dispatching model under a dynamic and stochastic environment. Carey and Crawford [13] formulated a train scheduling model on a network of busy complex stations and designed a series of heuristics for finding and resolving train conflicts so as to satisfy various operational constraints and objectives. Caimi et al. [14] constructed a resource-constrained multicommodity flow model for conflict-free train routing and scheduling. Chang et al. [15] built a multiobjective programming model for the optimal allocation of passenger train services on an intercity high-speed rail line without branches.

The above-mentioned studies provide useful methods for the optimization of train schedule for railway networks during a particular time period. Nonetheless, research to date has focused primarily on scheduling problem with even headways, and a unified framework for scheduling methods that can consider uneven headways and time-dependent demand patterns is critically limited. This paper focuses on an intercity rail line and proposes a phase-regular scheduling method, which divides a day equally into several time blocks and

applies the regular train-departing interval and the same train length during each period. An optimization model is presented to analytically calculate the passenger waiting time and the in-train crowded cost under a dynamic demand condition.

The remainder of this paper is organized as follows. An optimization model to the train scheduling problem for an intercity rail line is given in Section 2. In Section 3, a genetic algorithm procedure with two-layer framework is presented. In Section 4, there is a numerical example provided to illustrate the application of the proposed model and algorithm. The last section brings the paper to a conclusion and outlines the possibilities for future research in related areas.

2. Formulation

2.1. Problem Statement

This study considers the train operations along one direction at an intercity rail line which consists of K stations as shown in Figure 1. The stations along the direction are numbered as $1, 2, \dots, K$, and i is used to index a station. Assume that all trains have the same speed between two consecutive stations and R^i is used to denote the running time between station i and $i + 1$.

The passengers traveling by intercity rail lines arrive at the stations randomly, and they wait for the latest arriving train and then reach their destinations. The demands associated with intercity rail lines are characterized as being dynamic and stochastic. In terms of demand, the travel purposes of passengers are mainly for work and business, and the phase aggregation is significant. Based on such considerations, this paper divides a day equally into several time blocks (e.g., 1 hour as a period) and uses τ to denote a period and \mathfrak{R} to represent the set of periods ($\tau \in \mathfrak{R}$). At the same time, this study uses $D^{i,i+s}(\tau)$ to denote the number of passengers who arrive at station i during period τ travelling to station $i + s$.

Based on the characteristics of passenger demands and train operations about the intercity rail systems, a phase-regular train schedule, which has an even headway and the same train length during each period, is adopted in this study. According to the operation practice of intercity rail systems, this paper also assumes that there are only two train patterns, namely, the large pattern and the small pattern, in order to simplify the problem. The large pattern has more cars (e.g., 6 cars) and the small pattern has fewer cars (e.g., 4 cars) to form a train. To design a train schedule is actually to determine the number of departed trains and the associated train pattern for each period. The decision variables are defined as follows:

$x(\tau)$: number of scheduled trains during period τ ;

$y(\tau)$: binary variables indicating the train pattern during period τ , which equals to 1 if the large train pattern is adopted and 0 otherwise.

It is obvious that a train schedule for the intercity rail line can be transformed to calculate a set $\Omega = \{(x(\tau), y(\tau)) \mid \tau \in \mathfrak{R}\}$. In order to formulate a model accurately, two other hypotheses are presented as follows. At first, the travel passengers between two stations arrive uniformly at their origin station during a given period. Secondly, the passengers during a period are to be carried by the trains scheduled at the same period. The following notations

and parameters are defined for constructing a train scheduling model for the intercity rail line:

C_1 : train capacity with large pattern (e.g., 600 persons);

C_0 : train capacity with small pattern (e.g., 340 persons);

h_{\min} : prespecified minimum interval between two consecutive trains at the same station (e.g., 5 min);

h_{\max} : prespecified maximum interval between two consecutive trains at the same station (e.g., 30 min);

N_1 : number of provided train-units with large pattern at the origin station;

N_0 : number of provided train-units with small pattern at the origin station;

$Q_j^i(\tau)$: number of in-train passengers while train j departs from station i during period τ ;

$P_j^{i+s}(\tau)$: number of passengers boarded on train j who arrive at station i travelling to station $i + s$ during period τ .

2.2. Constraints

(1) *Train Operation Constraint.* The minimum interval between two consecutive trains should be required to ensure the operation safety of trains, while the predetermined maximum interval should not be broken for the passenger waiting times at the stations cannot be too long. Considering that the number of trains $x(\tau)$ is scheduled during period τ , the same headway between two consecutive trains during this period is thus denoted by $60/x(\tau)$ (min). As a result, the train operation constraint can be expressed by the following inequality:

$$h_{\min} \leq \frac{60}{x(\tau)} \leq h_{\max}. \quad (2.1)$$

(2) *Demand and Supply Constraint.* According to the second hypothesis, the passenger demands generated during a period should be fulfilled by the trains scheduled at the same period. The demand and supply constraint is determined by the following:

$$\sum_{j=1}^{x(\tau)} P_j^{i,i+s}(\tau) = D^{i,i+s}(\tau). \quad (2.2)$$

(3) *The In-Train Passengers.* When train j departs from station i during period τ , the number of in-train passengers contains two parts: one is the boarded passengers whose destinations are farther than station i , and the other is the passengers boarding at station i . Thus, the in-train passengers can be calculated as follows.

$$Q_j^i(\tau) = \sum_{i_1=1}^i \sum_{i_2=i_1+1}^K P_j^{i_1,i_2}(\tau). \quad (2.3)$$

(4) *The Fleet-Size Constraint.* For the intercity rail transit lines with high-frequency train schedules, it is very important to have enough train units for dispatching at any moment. This paper considers that the fleet size is the major resource constraint in our scheduling design problem. Considering that the number of train units available at the origin station is known in advance, the fleet size constraint can be expressed as follows:

$$\sum_{\tau \in \mathcal{R}} x(\tau) \cdot [k \cdot y(\tau) + (1 - k) \cdot (1 - y(\tau))] \leq k \cdot N_k + (1 - k) \cdot N_{1-k} \quad (k = 1, 0). \quad (2.4)$$

(5) *The Number of Boarded Passengers.* The number of boarded passengers traveling from station i to station $i + s$ for train j during period τ , $P_j^{i,i+s}(\tau)$, is actually to assign the passengers to different trains. Considering that the passengers arrive uniformly at the original station and the trains associated with a given period are scheduled with a constant headway, the passenger demands should be evenly assigned to the trains during the period. The following formula is thus achieved for calculating the number of boarded passengers:

$$P_j^{i,i+s}(\tau) = \begin{cases} \left\lceil \frac{D^{i,i+s}(\tau)}{x(\tau)} \right\rceil & \text{if } 1 \leq j \leq x(\tau) - R \\ \left\lceil \frac{D^{i,i+s}(\tau)}{x(\tau)} \right\rceil + 1 & \text{if } x(\tau) - R < j \leq x(\tau), \end{cases} \quad (2.5)$$

where $\lceil \cdot \rceil$ means rounding, $D^{i,i+s}(\tau) = \lceil D^{i,i+s}(\tau) / x(\tau) \rceil \cdot x(\tau) + R$, $0 \leq R \leq x(\tau) - 1$.

2.3. Objective Function

The objective function is to minimize the total costs, which are composed of the waiting times of passengers at stations and the in-train crowded costs.

A constant interval between two consecutive trains is $60/x(\tau)$ during period τ for there are $x(\tau)$ trains operating at this period. If the passengers arrive at station i with uniform distribution, the average waiting time of each passenger during the period is $30/x(\tau)$. The number of passengers arriving at station i during the period is $\sum_{s=1}^{K-i} D^{i,i+s}(\tau)$, and the total waiting times of passengers at station i is $(30/x(\tau)) \cdot \sum_{s=1}^{K-i} D^{i,i+s}(\tau)$. Thus, the total waiting times of passengers during period τ can be expressed as follows:

$$W(\tau) = \sum_{i=1}^{K-1} \sum_{s=1}^{K-i} \frac{30}{x(\tau)} \cdot D^{i,i+s}(\tau). \quad (2.6)$$

This study introduces the in-train crowded cost to evaluate travel condition in the trains. The cost is incurred while the number of onboard passengers exceeds the maximum loading capacity of a train. As a result, the in-train crowded cost of train j running between station i and station $i + 1$ during period τ is $Q_j^i(\tau) \cdot R^i$ if

$$Q_j^i(\tau) > \begin{cases} C_1 & \text{if } y(\tau) = 1 \\ C_0 & \text{if } y(\tau) = 0 \end{cases} \quad (2.7)$$

and 0 otherwise. The total in-train crowded cost during period τ , $F(\tau)$, is thus presented as follows:

$$F(\tau) = \sum_{j=1}^{x(\tau)K-1} \sum_{i=1} Q_j^i(\tau) \cdot R^i \cdot \delta \left[Q_j^i(\tau) - (y(\tau) \cdot C_1 + (1 - y(\tau)) \cdot C_0) \right], \quad (2.8)$$

where $\delta(u)$ is the sign function which is equal to 1 if $u > 0$ and 0 otherwise.

According to the above discussions, the objective function for minimizing the waiting times of passengers at stations and the in-train crowded costs can be expressed as follows:

$$\min Z = \sum_{\tau \in \mathcal{R}} [F(\tau) + W(\tau)]. \quad (2.9)$$

3. Solution Algorithm

The proposed model is a nonlinear programming problem which associates the tightly related zero-one and integer variables. It can hardly be solved with conventional gradient-based methods or commercial optimization solvers. Based on the mechanics of natural selection and natural genetics, genetic algorithm [16–18] is therefore adopted to solve the model developed in this study. The appeal of genetic algorithm comes from its simplicity and elegance as robust search algorithm as well as from its power to discover good solutions rapidly for difficult high-dimensional problems. In particular, a hybrid procedure with two-layer framework is designed to solve the proposed model.

3.1. Encoding Approach

A special coding approach with two-layer structure, which includes two layers of decision variables, namely, the number of trains and train pattern, is adopted to solve the model. A day is divided equally into several short periods (e.g., one hour as a period in this study), and the length of a chromosome is represented by the number of periods. An integer embedded with the upper layer encoding indicates the number of trains scheduled during the given period. A binary number embedded with the lower layer encoding means the train pattern, where 1 indicates the large pattern and 0 indicates the small pattern.

In view of the headway constraint, the value of integers associated with the upper layer encoding, and the number of scheduled trains, should be located within $[60/I_{\max}, 60/I_{\min}]$. The number of scheduled trains $x(\tau)$ should range from 2 to 12, for example, if $I_{\min} = 5$ min and $I_{\max} = 30$ min. Figure 2 illustrates a chromosome using two-layer coding approach, where a day is divided into 10 periods, the number of scheduled trains is 6, 11, 7, 6, 9, 5, 4, 2, 8, 5, and the train pattern is large, large, small, small, large, large, small, small, large, small, respectively, during the concerned period. Following the above-mentioned method, we can generate the initial chromosomes at random, then determine the number of trains $x(\tau)$ and the train pattern $y(\tau)$ and, finally, calculate the relevant parameters and the objective function.

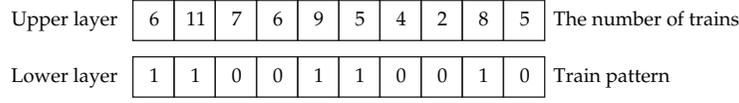


Figure 2: The illustration of a chromosome with two-layer coding approach.

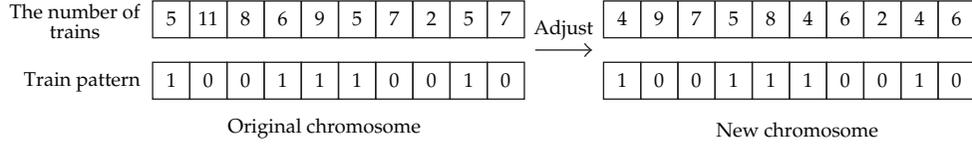


Figure 3: The adjustment of an infeasible chromosome.

3.2. Feasibility Adjustment

It is necessary to pay special attention to the fleet size constraint in the proposed model, because the number of trains departing from the original station during one day should not exceed the provided train units, and the chromosomes are required to adjust the feasibility as the execution process of the procedure. The total number of trains with large pattern is calculated by checking the train pattern whose encoding is 1 at the lower layer. The number of trains $x(\tau)$ associated with period τ should be reduced by 1 with descending order if the total number of trains exceeds the available train units N_1 , until the condition is satisfied. The total number of trains with small pattern is required to check its feasibility similarly. All other chromosomes generated at the iterative process of genetic operation should be also adjusted with the above-mentioned method. For example, a new chromosome as shown in Figure 3 can be obtained after the adjustment of the original one if the number of available train units N_1 and N_0 are 25 and 30, respectively.

3.3. Fitness Function

By calculating the value of objective function, we can get the fitness of each chromosome. The formulation can be expressed as follows:

$$\text{fitness} = \frac{Z_{\max} - Z}{Z_{\max} - Z_{\min}}, \tag{3.1}$$

where Z is the objective from (2.9), Z_{\max} and Z_{\min} denote, respectively, the maximum and minimum values of the objectives associated with the current generation.

3.4. Crossover Operator

Considering that two-layer encoding approach is adopted in this paper, a layered crossover operation with double probabilities is proposed. The procedure of the layered crossover operation is presented as follows.

Algorithm 3.1. The layered crossover operation consists of the following steps.

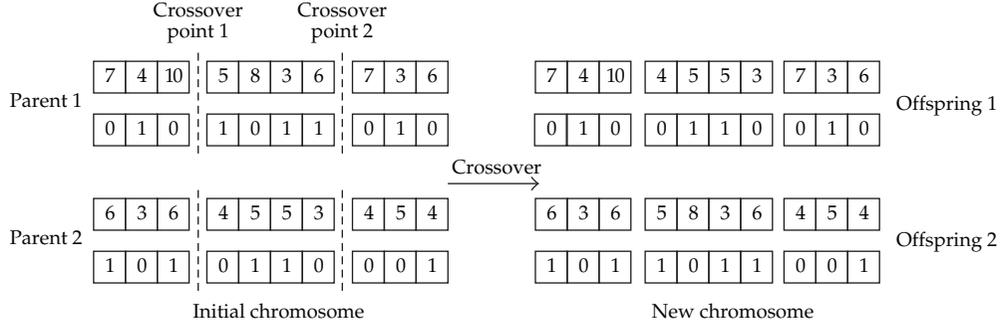


Figure 4: The illustration of two-layer crossover operation.

Step 1. Two crossover probabilities P_c^1 and P_c^2 are set to decide the upper layer operation and the lower layer operation, and a random number θ with uniform distribution at $[0, 1]$ is generated to indicate the judgment criterion.

Step 2. Two crossover points are selected randomly, then two gene strings between two crossover points on the parent chromosomes are exchanged with each other.

Step 3. Two new offspring chromosomes are generated by exchanging the corresponding gene strings of the parent chromosomes if $\theta \leq \min\{P_c^1, P_c^2\}$, which means that two gene strings associated with the upper layer and the lower layer are exchanged simultaneously between parent 1 and parent 2. If $P_c^2 < \theta \leq P_c^1$, only the gene strings associated with the upper layer of two parents take the crossover operation, and the gene strings associated with the lower layer remain the same. If $P_c^1 < \theta \leq P_c^2$, only the gene strings associated with the lower layer of two parents take the crossover operation, and the gene strings associated with the upper layer remain the same.

Assume that the crossover probability $\theta \leq \min\{P_c^1, P_c^2\}$, the layered crossover operation can be demonstrated by Figure 4.

3.5. Mutation Operator

According to the characteristics of the problem, the large train pattern is suitable for the period with larger number of trains, while the small train pattern is suitable for the period with less number of trains. A layered mutation operation with a single point is then proposed in the paper. Firstly, a gene associated with the upper layer encoding takes mutation operation, and the corresponding gene with the lower layer is then operated by the result of the upper layer. The procedure of the layered mutation operation is summarized as follows.

Algorithm 3.2. The layered mutation operation consists of the following steps.

Step 1. Two mutation probabilities P_m^1 and P_m^2 are set to decide the mutation operations associated with the upper layer and the lower layer, respectively, denoting the judgment criteria of the mutation operation with upper and lower layers, and a random number θ with uniform distribution at $[0, 1]$ is generated to indicate the judgment criterion.

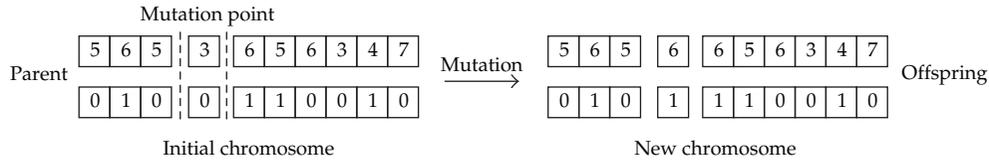


Figure 5: The illustration of the layered mutation operation.

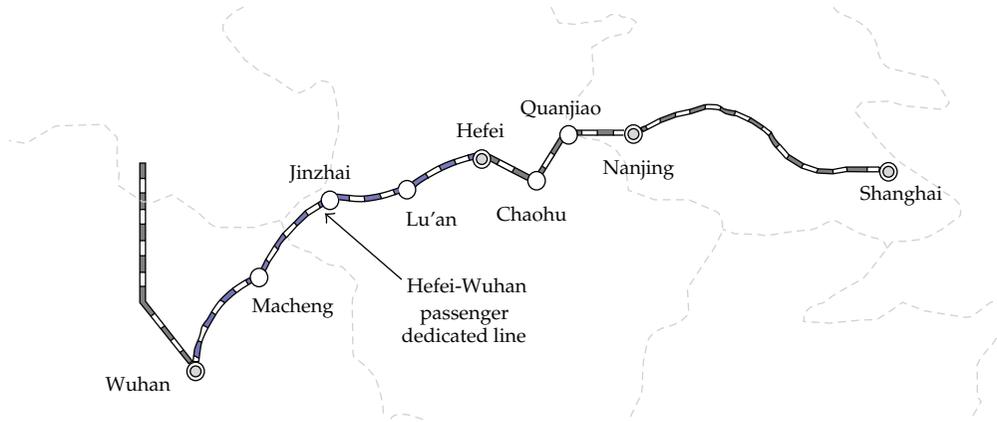


Figure 6: The illustration of Hefei-Wuhan intercity rail line.

Step 2. A mutation point is selected randomly, and then the genes associated with the upper layer and lower layer are mutated.

Step 3. The selected gene at the upper layer of the parent chromosome is replaced with another number which is located within the predetermined range if $\theta \leq P_m^1$. The mutation operation with the lower layer will be determined by the result of the upper layer if $\theta \leq P_m^2$. If the value of the gene at the upper layer encoding increases, the value of the corresponding gene at the lower layer should be changed from 0 to 1. If the value of the gene at the upper layer decreases, the value of the corresponding gene at the lower layer encoding should replace 1 with 0.

Assume, for example, that the mutation probability is $\theta \leq P_m^1$ and $\theta \leq P_m^2$, the layered mutation operation can be demonstrated by Figure 5.

4. Numerical Example

4.1. Line

Hefei-Wuhan intercity rail line, which has operated since April, 2009, is an important passenger dedicated railway line between Hefei city and Wuhan city in China. The line has a total length of 364 kilometers with the designed speed of 250 km/h. There are 5 stations along the line, namely Wuhan station, Macheng North station, Jinzhai station, LiuAn station, and Hefei station as shown in Figure 6.

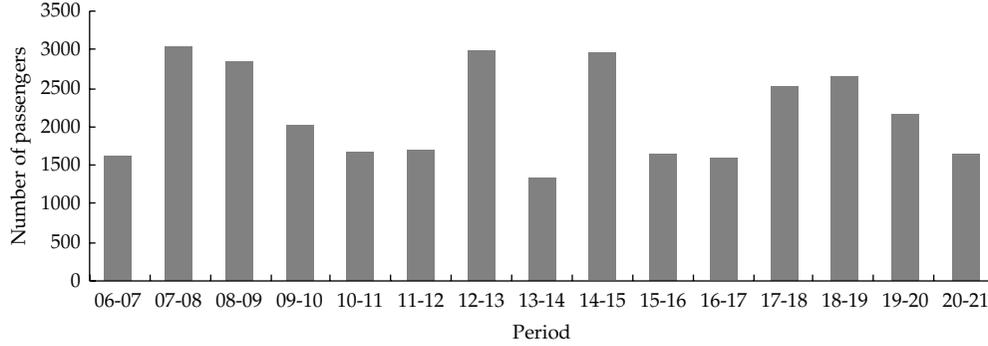


Figure 7: The period-dependent passenger demands counted in terms of one hour.

Table 1: The running time between two adjacent stations (min).

Adjacent stations	Running time	Adjacent stations	Running time
Wuhan-Macheng North	52	Jinzhai-Liuan	22
Macheng North-Jinzhai	44	Liuan-Hefei	32

4.2. Demands

With a particular focus on a typical weekday, the example below considers the operation from 6:00 AM to 21:00 PM at the origin station. The period-dependent passenger demands, as shown in Figure 7, are illustrated by the total number of passengers arriving at the stations during each one hour.

4.3. Results

The running times between two adjacent stations for Hefei-Wuhan intercity rail line are given in Table 1. The capacities associated with two train patterns C_1 and C_0 are 600 and 340, respectively, and the fleet sizes of available train units N_1 and N_0 at the origin station are 40 and 50, respectively. The prespecified minimum and maximum headways I_{\min} and I_{\max} are 5 and 30 minutes, respectively.

The parameter values for the layered genetic algorithm are listed as follows. The population size is 100 and the number of total iterations is 300. The crossover probabilities P_c^1 and P_c^2 are 0.90 and 0.95, and the mutation probabilities P_m^1 and P_m^2 are 0.10 and 0.30, respectively. After 252 iterations, a train schedule for Hefei-Wuhan intercity rail line from 6:00 AM to 21:00 PM is obtained by the developed procedure, which is shown in Table 2, and the trend of objective value in processing the algorithm can be shown in Figure 8.

From Table 2, we can see that the total number of trains scheduled during the operation period is 77, of which the number of trains with large pattern is 37, and the number of trains with small pattern is 40. The average full-load rate of trains is 95.65%, which means the optimized train schedule could both economize the operation cost for the railway department and provide comfortable travel environment for the passengers. The number of scheduled trains and the train pattern during each period are shown in Figure 9.

We can see from Figure 9 that the number of scheduled trains is generally proportional to passenger demand, which means that the number of trains during high-peak periods is

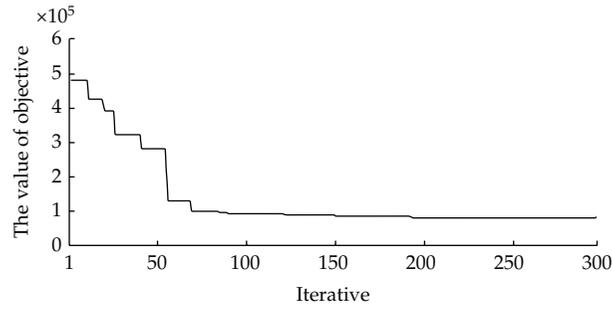


Figure 8: The variation curve on the objective with iterations.

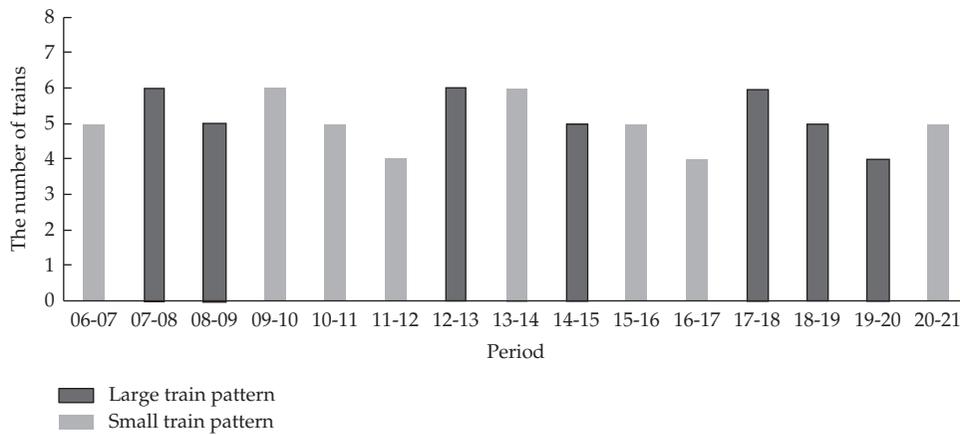


Figure 9: The number of scheduled trains and the train-pattern during each period.

Table 2: The number of trains and train pattern during each period.

Period	Number of trains	Train pattern	Average load-rate (%)
6:00-7:00	5	0	96.00
7:00-8:00	6	1	85.14
8:00-9:00	5	1	95.50
9:00-10:00	6	0	99.71
10:00-11:00	5	0	99.24
11:00-12:00	4	0	99.88
12:00-13:00	6	1	83.64
13:00-14:00	6	0	90.69
14:00-15:00	5	1	99.57
15:00-16:00	5	0	97.53
16:00-17:00	4	0	94.71
17:00-18:00	6	1	70.58
18:00-19:00	5	1	89.57
19:00-20:00	4	1	90.21
20:00-21:00	5	0	97.53

larger and vice versa. The train pattern is mainly with large pattern during the high-peak periods, and during the low-peak periods the train pattern is mainly with small pattern.

5. Conclusion

This paper proposes a phase-regular scheduling method for an intercity rail line, which divides an operational day evenly into several time blocks and applies a regular train-departing interval and the same train length for each period. A nonlinear mixed zero-one programming model, which could accurately calculate the passenger waiting time and the in-train crowded cost, is established. A hybrid genetic algorithm with two-layer framework is designed to solve the proposed model. Finally, the validation of the model and the algorithm has been tested with the application of Hefei-Wuhan intercity rail line in China. The results show that the proposed method can effectively solve the scheduling problem of intercity rail lines. Considering the modeling details which are closer to reality, such as under a random or fuzzy environment, is an important topic for further research. At the same time, there is the necessity to explore the response of passengers to the optimized schedule and to extend the method to a network case.

Acknowledgments

The work described in the paper was supported by National Natural Science Foundation of China (no. 50968009, no. 71261014, and no. 51210305046) and the Research Fund for the Doctoral Program of Higher Education (no. 20096204110003).

References

- [1] N. S. A. Ghoneim and S. C. Wirasinghe, "Optimum zone structure during peak periods for existing urban rail lines," *Transportation Research B*, vol. 20, no. 1, pp. 7–18, 1986.
- [2] J. W. Goossens, S. van Hoesel, and L. Kroon, "On solving multi-type railway line planning problems," *European Journal of Operational Research*, vol. 168, no. 2, pp. 403–424, 2006.
- [3] C. Liebchen, "The first optimized railway timetable in practice," *Transportation Science*, vol. 42, no. 4, pp. 420–435, 2008.
- [4] M. T. Claessens, N. M. Van Dijk, and P. J. Zwaneveld, "Cost optimal allocation of rail passenger lines," *European Journal of Operational Research*, vol. 110, no. 3, pp. 474–489, 1998.
- [5] 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.
- [6] M. B. Khan and X. Zhou, "Stochastic optimization model and solution algorithm for robust double-track train-timetabling problem," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 81–89, 2010.
- [7] X. Zhou and M. Zhong, "Single-track train timetabling with guaranteed optimality: branch-and-bound algorithms with enhanced lower bounds," *Transportation Research B*, vol. 41, no. 3, pp. 320–341, 2007.
- [8] K. Nachtigall and S. Voget, "Minimizing waiting times in integrated fixed interval timetables by upgrading railway tracks," *European Journal of Operational Research*, vol. 103, no. 3, pp. 610–627, 1997.
- [9] A. de Palma and R. Lindsey, "Optimal timetables for public transportation," *Transportation Research B*, vol. 35, no. 8, pp. 789–813, 2001.
- [10] S. Nguyen, S. Pallottino, and F. Malucelli, "A modeling framework for passenger assignment on a transport network with timetables," *Transportation Science*, vol. 35, no. 3, pp. 238–249, 2001.
- [11] R. C. W. Wong, T. W. Y. Yuen, K. W. Fung, and J. M. Y. Leung, "Optimizing timetable synchronization for rail mass transit," *Transportation Science*, vol. 42, no. 1, pp. 57–69, 2008.

- [12] L. Meng and X. Zhou, "Robust single-track train dispatching model under a dynamic and stochastic environment: a scenario-based rolling horizon solution approach," *Transportation Research B*, vol. 45, no. 7, pp. 1080–1102, 2011.
- [13] M. Carey and I. Crawford, "Scheduling trains on a network of busy complex stations," *Transportation Research B*, vol. 41, no. 2, pp. 159–178, 2007.
- [14] G. Caimi, F. Chudak, M. Fuchsberger, M. Laumanns, and R. Zenklusen, "A new resource-constrained multicommodity flow model for conflict-free train routing and scheduling," *Transportation Science*, vol. 45, no. 2, pp. 212–227, 2011.
- [15] 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.
- [16] M. Gen and R. W. Cheng, *Genetic Algorithms and Engineering Optimization*, John Wiley & Son, New York, NY, USA, 2000.
- [17] H. M. Niu, "Determination of the skip-stop scheduling for a congested transit line by bilevel genetic algorithm," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1158–1167, 2011.
- [18] J. Gao, R. Chen, and Q. Pan, "A hybrid genetic algorithm for the distributed permutation flowshop scheduling problem," *International Journal of Computational Intelligence Systems*, vol. 4, no. 4, pp. 497–508, 2011.

Research Article

Coherent Network Optimizing of Rail-Based Urban Mass Transit

Ying Zhang

School of Transportation Engineering, Tongji University, 4800 Caoan Highway, Shanghai 201804, China

Correspondence should be addressed to Ying Zhang, yzzy@tongji.edu.cn

Received 7 July 2012; Revised 10 September 2012; Accepted 26 September 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Ying Zhang. 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.

An efficient public transport is more than ever a crucial factor when it comes to the quality of life and competitiveness of many cities and regions in Asia. In recent years, the rail-based urban mass transit has been regarded as one of the key means to overcoming the great challenges in Chinese megacities. The purpose of this study is going to develop a coherent network optimizing for rail-based urban mass transit to find the best alternatives for the user and to demonstrate how to meet sustainable development needs and to match the enormous capacity requirements simultaneously. This paper presents an introduction to the current situation of the important lines, and transfer points in the metro system Shanghai. The insufficient aspects are analyzed and evaluated; while the optimizing ideas and measurements are developed and concreted. A group of examples are used to illustrate the approach. The whole study could be used for the latest reference for other megacities which have to be confronted with the similar situations and processes with enormous dynamic travel and transport demands.

1. Introduction

An efficient public transport is more than ever a crucial factor when it comes to the quality of life and competitiveness of many cities and regions. Rail-based urban mass transit is the backbone of modern transit systems [1], as their inherent characteristics such as exclusive right-of-way, automated guidance, and electric propulsion allow them to carry large numbers of passengers with speed, convenience, and safety.

Projects of rail-based urban mass transit are often a top contender to meet the rapidly increasing travel demand, especially in many Asian cities [2]. As high population density requires an efficient transport system to facilitate mobility and economic development of the territory. On the other hand, this high density provides the essential condition for the development of a mass transit type of public transport. Furthermore, this kind of transit system is bound to grow. The world is constantly urbanizing, creating megacities that rely more and more on public transportation.

Indeed, the development of rail-based urban mass transit has been one of the key objectives of the transport policies and strategies over the past few decades. As the population size of Chinese city increases, the use of a rail-based system becomes more appealing. Since 2003, 28 cities have done their plans of this system network. 2700 km among the planned lines have been constructed after that. This means, above 250 km line track per year have been brought into operation, which is quite faster than the construction speed of in western world in 70's last century with 160 km per year.

Despite the recent remarkable achievements, the development of rail-based urban mass transit system in China still faces a lot of problems [3]. Especially, this kind of remarkable development causes insufficient aspects, such as deficiency and inconsistency. Some plans have to be changed. The other must even go through fundamental revision.

Transit networks, although being structurally simpler than many other networks, present some specific challenges. The network planning of rail-based urban mass transit is influenced and restricted by a series of factors, such as nature environment, social system, economic condition, and construction technique, so that its concepts are normally multifarious with different emphasizes [4].

First of all, there are a number of surveys of the empirical literature on demand and capacity of network. There is, therefore, no contribution to be made by providing an extensive literature survey here.

Secondly is about the evaluation of network planning, which is difficult to measure, and while there is no consensus within the literature on a single appropriate proxy variable, three approaches could be defined that are most commonly used: (1) number of vehicles/trains in operation and train kilometers operated [5] (as these figures increase per passenger or per route length, service frequency also increases and there is less crowding), (2) some measure of time or money [6] (for instance in vehicle time or waiting time through a value of time factor), and (3) other quality factors that are not directly measurable in terms of time or money [7] (such as service reliability, infrastructure quality, and ventilation).

Some approaches have also used simulation models that are constructed to describe the relationship [8]. Normally the simulation model is a generic model that can be changed to adapt the influencing factors, such as rate of carriage fullness or length of time periods [9]. By running the model, several what-if questions can be replied to make revisions.

Finally, the importance of intermodal connectivity at stations [10], which had been argued and was also confirmed by many studies. Providing good feeder bus services in both origin and destination stations at all time periods considerably enhanced ridership [11]. At the same time, the walk ability for pedestrians in station proved to be important for increasing ridership.

But analysis in a multimodal network is more complicated than the investigation of pure vehicular or bus trips. It involves combinedmode trips in which travelers choose not only the routes, but also the transport modes and the kinds and locations of transfers [12]. Some studies have demonstrated that combining this framework with sensitivity and game theoretic approaches form a platform for analyzing the competition between operators as well as for studying the case of regulation [13, 14]. Furthermore, considering high demand for metro systems, a service disruption may lead to significant degradations in a city's public transportation system. That is why establishing alternative means of transportation for passengers usually using buses is undertaken by transport authorities [15].

While traditional transit planning methods consider such characteristics as demand, demography, geography, time, and others, none seem to address the network design in a direct manner, which becomes increasingly important as systems grow. Till now, the most

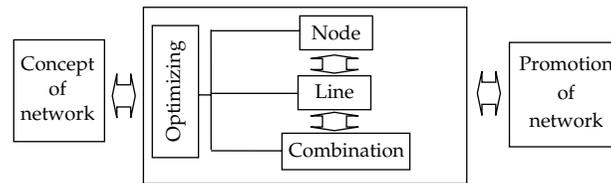


Figure 1: Coherent constituent parts of optimizing.

applied optimizing methods have been struggling to include too many influences and variants. Lacking of an efficient and effect optimizing system for the network planning should be the one of the main reasons of this dilemma situation. This paper could shed light on it.

2. Efficient and Effect Optimizing

According to the philosophy of planning, the character of process should be emphasized, especially during the phase of planning revelation. As conventional assessment approach normally focuses on picking out the best one among all candidate concepts at the last phase of network layout, while an efficient and effect optimizing contrarily attends to value every main elements of network during the whole planning step by step.

It is worthy to mention, the principle of this coherent network optimizing is to achieve the biggest accumulated net contribution, which is equal to total profit mines total cost. Profit results from ticket and increment of land value, while cost will be calculated not only with construction investment, but also operation expense as well as passenger mobility time.

This optimizing consists of main three constituent parts; they are node and line as well as combination of network (see Figure 1). As mentioned before, this process should be infiltrated into every course of network planning, so that planner could benefit through planning coupling to promote the quality of every concept and finally to achieve more valuable and sustainable network plan in respects of different aspects of point, line, and surface.

2.1. Optimizing of Main Node

It is crucial in convenience and possibility of transfer, as node plays an important roll to form the whole network. It decides the level of accessibility and flexibility for whole network. At the same time, it also influences the cost and construction tempo, especially when they are directly located in the central city.

Because of the limited space and density buildings in these areas, potential space is always limited. Besides the above necessity analyses, its layout and construction cost should be good investigated in each concept which can be multiplied the sum of prognostic passenger by its unit price of mobility time.

2.2. Optimizing of Line

It determines the form and difficulty level of network. The terrace of line affects mostly the construction feasibility, so that it decides the correlative construction cost finally. The needs of operation, such as park place of wagon maintenance place, and, should be also considered by the choice of route.

If the line needs to be taken by special construction measure, its cost will be raised with several times. For this reason, it is worthy through quantitative analysis of the demand about space, technique, and expense during the planning carefully.

2.3. Optimizing of Combination

Combination means the mechanics of network. It overlaps with the front two aspects and covers the capacity and capability of the transit network. According to the transfer times and its time cost, the final layout of different lines, either crossing or parallel, can be set up through breaking, remarking, and interweaving.

Nevertheless, a group of operation index such as passenger flow, volume of circular flow, and rate of flow in different directions can be regarded as criteria to argue the efficiency of network operation.

Capacity is a quite vital important for network planning, as it refers to the adaptability of the present and future transport volumes. Absolutely it is the biggest challenge rising in planning of rail-based urban mass transit in China, where the urbanization, modernization, and motorization generate rapidly and enormously. These kinds of dramatic development bring gigantic demand and stress of passenger flow to the rail-based urban mass transit, so that they are definitively among the main reasons of deficiency and insufficiency of network planning mentioned at the beginning of this paper.

3. Application of Coherent Optimizing and Measurement

In order to give prominence to these mentioned aspects, some examples and cases shall be written in detail as follow.

As one of the Chinese megacities, public transport Shanghai has been challenged by overload of passenger volume day after day. According the last population census at the end of 2010, there are more than 2.3 million inhabitants in Shanghai. Obviously, it exceeds the prediction sum of population 2020 stated in the master plan with the total amount of 1.6 million, which was however quoted as basic reference date for the plan of whole system.

So it is not difficult to understand, why the lines are quite overloaded and crowded during peak hours, and why the network can still not meet the huge demand even after rapid construction in last decade years.

By means of the elevated ring streets, the network of the rail-based urban mass transit Shanghai can be divided into three parts (see Figure 2). Part one is the core area which covers the city center with about a radius of 3 km. In this area, there are 64 km lines and 34 metro stations. Secondly, part two is the area between core area and inner ring, which circular is the rest part of the central city without above mentioned core area. Here are located altogether 208 km lines and 108 stations. Finally, part three is the third one of network, which is outside of the inner ring street and stretches in the city periphery.

In this fall, the service radius of network is, respectively, 500 m in core area and 600 m in central area, which is closer and longer than the normal recommended value in other world cities. However, the service radius directly affected the mobility time and accessibility as well as the efficiency of whole network.

So as a measurement is called, the density of node and lines inside both of the core area and central area should be further completed. Especially near the present huge hubs, which are often overloaded, could be through addition of stations to ease the satiation degree.

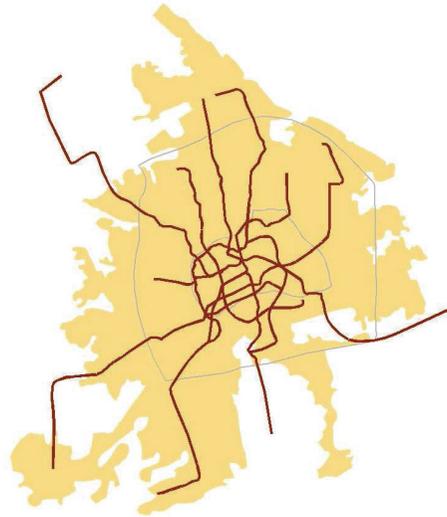


Figure 2: Shanghai rail-based urban mass transit (in red) and inner- and outside ring streets (in grey) as well as urban constructed area (in yellow).

Example 3.1 (transfer of node). In the station of People's Square, where is the super center of administration, culture, and business in Shanghai, there are 3 metro lines concentrated under the ground, in addition with more than 40 bus lines above the ground (see Figure 3). The station of People's Square is in the mitten crossing of line 1, 2, 8. All of them combine the biggest traffic hub for the urban and region. The volume of passenger and transfer passenger altogether accounts for 900 thousand per day, which is quantitatively equal to the dimension of population in a big city.

With the urbanization and development of real estate in suburb areas, more and more people remove from central city to outskirts of the city. At the same time, the central city has won more and more area to be redeveloped and reconstructed. Therefore, the connection between central city and suburb has become more important than before.

During peak hour, most of this kind of connection has relied on rail transit system, especially the radial pattern lines. Although, this phenomenon is accord with the classic economic geography theory, but it is also confronted with typical Chinese feature, namely, masses of passengers. One reason for that is over speeding sprawl of the city and excessive development of real estate around rail lines.

Example 3.2 (radial pattern of line). Since 1995, the line 1 goes from south-west via center to north as first operated line in Shanghai and binds new developing zone via developed city center to the main railway station. With the development of rail-based urban mass transit network, the average rate of growth of passenger volume in each year is about 16% from 2001 to 2011, after the other main lines have also been operated in business inside of city area.

But the growth rates for every station along line 1 are quite different. In south-west are more than 50%, because of density development of new residential areas. These growth rates with value higher than 20% are originated near the new shopping center or new business center, which are located between city center and new residential area. Most stations have experienced growth about 10%–20% rate. Few of them have been increased less than 5%. And very few are even gone down (see Figure 3).

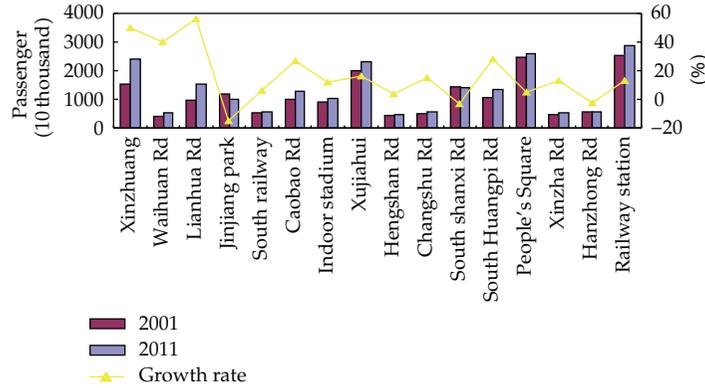


Figure 3: Growth rate of passenger flow of line 1.

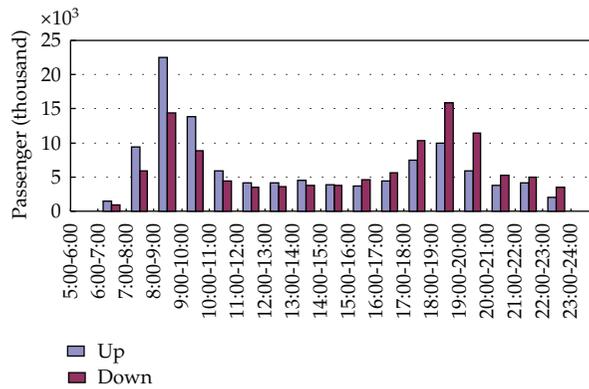


Figure 4: Daily Passenger flow in two ride directions of line 8 in 2010.

For another example, line 8 is overloaded with rapid growth rate because of the connection of large scale of residential communities in two ends. During peak hour at normal day, some sections of this line have to withstand the full-load ratio 175% (see Figure 4).

Example 3.3 (seamless link between city and region). Another main reason goes to the restriction of network density in outside of the city. There are only 8 lines among 18 radial pattern lines, which extend to the distance longer than 30 km. The last stations of other 10 lines break down less than 20 km outward from the city center (see Figure 2).

Incidentally, the distance with 30 km is about the bound line by rail-based urban mass transit system for commuter traffic. With the normal speed of 30–40 km per hour, rail-based urban mass transit can be as the best suitable traffic modal to give passenger about 1 hour trip within this region.

If the capacity of commuter traffic between city center and suburb could not keep the same step with the increase demand, its advantages, such as express, reliable, and massive, would not be exerted. Furthermore, it would result in more and more private motorized trip and would bring worse unhealthy modal split.

Tracing to the development of rail transit system in advanced world cities, the measurement is strongly recommended to add long radial pattern lines with 30 km distance at least to 20 lines. In that fall, it can achieve 2 lines in same direction for every periphery area, and avoid the service problem rising from the interval space between two radial pattern lines.

The third measurement to increase the capacity and capability can be realized by establishment of regional express line. Regional express line distinguishes rail-based urban mass transit with its relative longer interval distance between stations, faster speed and bigger capacity. When it is introduced to the network, it can improve the long connection between central city and periphery as passenger corridor.

At the same time, it can also reduce the almost to satiation volume pressure inside of the central area. Moreover, it can promote the impact power of Shanghai in the Yangtze delta area, so that it is quite reasonable to be built for not only for the reason of transport, but also for the future of region development.

4. Conclusions

The coherent network optimizing is useful to promote the quality of network plan. The analysis of four aspects, namely, node, route, formation, and capacity conduce to assess each concept comprehensively and abjectly, especially when the principle of efficiency about net contribution is in every course compared.

Last but not least, it is also, facility to be applied in comparison not only for the whole of network, but also for some sectors. These examples in Shanghai have shown, that both of central area and periphery area should be emphasized in network planning, especially during the phase of urbanization.

And urbanization has been a worldwide phenomenon since last century. By 2050, above 70% of the world population will live in urban areas, although the figure is right now about the half of them. Shanghai is the typical example in China, which has made fast pace of urbanization particularly and continually for the past three decades.

All of urban growth, urban sprawl, and increased motorization have brought significantly the demand of transport and require major adaptation of the urban infrastructures, in particular public transport. In this context, an important goal has to be made to reach a high modal share of public transport requires. This study offers an initial step toward that direction.

It has been demonstrated, that rail-based urban mass transit can be as the backbone of public transport to meet the actual mobility needs and to inform the choices ahead for future urban mobility. The promotion of its network planning through coherent network optimizing will be useful to capitalize on mutual benefits. So, it could be used as the latest reference for other megacities which have to be confronted with the similar situations and processes with enormous travel and transport demands.

References

- [1] V. Vuchic, *Urban Transit: Operations, Planning and Economics*, John Wiley & Sons, Hoboken, NJ, USA, 2004.
- [2] S. Tang and H. K. Lo, "The impact of public transport policy on the viability and sustainability of mass railway transit—the Hong Kong experience," *Transportation Research Part A*, vol. 42, no. 4, pp. 563–576, 2008.

- [3] B. P. Y. Loo and D. Y. N. Li, "Developing metro systems in the People's Republic of China: policy and gaps," *Transportation*, vol. 33, no. 2, pp. 115–132, 2006.
- [4] M. Gunduz, L. O. Ugur, and E. Ozturk, "Parametric cost estimation system for light rail transit and metro trackworks," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2873–2877, 2011.
- [5] D. J. Graham, A. Crotte, and R. J. Anderson, "A dynamic panel analysis of urban metro demand," *Transportation Research Part E*, vol. 45, no. 5, pp. 787–794, 2009.
- [6] V. Guihaire and J. K. Hao, "Transit network design and scheduling: a global review," *Transportation Research Part A*, vol. 42, no. 10, pp. 1251–1273, 2008.
- [7] Y. Shafahi and A. Khani, "A practical model for transfer optimization in a transit network: model formulations and solutions," *Transportation Research Part A*, vol. 44, no. 6, pp. 377–389, 2010.
- [8] Ö. Yalçinkaya and G. Mirac Bayhan, "Modelling and optimization of average travel time for a metro line by simulation and response surface methodology," *European Journal of Operational Research*, vol. 196, no. 1, pp. 225–233, 2009.
- [9] S. Hassan, M. Salgado, and J. Pavon, "Friendship dynamics: modelling social relationships through a fuzzy agent-based simulation," *Discrete Dynamics in Nature and Society*, vol. 2011, Article ID 765640, 19 pages, 2011.
- [10] European Commission, "Group for urban interchanges development and evaluation (GUIDE)," in *The Fourth Framework Research and Technological*, 2000.
- [11] J. Y. Choi, Y. J. Lee, T. W. Kim, and K. Sohn, "An analysis of metro ridership at the station-to-station level in Seoul," *Transportation*, vol. 39, no. 3, pp. 705–722, 2012.
- [12] M. Smart, M. A. Miller, and B. D. Taylor, "Transit stops and stations: transit managers' perspectives on evaluating performance," *Journal of Public Transportation*, vol. 12, no. 1, pp. 59–77, 2009.
- [13] H. K. Lo, C. W. Yip, and Q. K. Wan, "Modeling competitive multi-modal transit services: a nested logit approach," *Transportation Research Part C*, vol. 12, no. 3-4, pp. 251–272, 2004.
- [14] D. Villatoro, S. Sen, and J. Sabater-Mir, "Of social norms and sanctioning: a game theoretical overview," *International Journal of Agent Technologies and Systems*, vol. 2, no. 1, pp. 1–15, 2010.
- [15] K. Kepaptsoglou and M. G. Karlaftis, "The bus bridging problem in metro operations: conceptual framework, models and algorithms," *Public Transport*, vol. 1, no. 4, pp. 275–297, 2009.

Research Article

Study on the Train Operation Optimization of Passenger Dedicated Lines Based on Satisfaction

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 22 August 2012; Accepted 13 September 2012

Academic Editor: Wuhong Wang

Copyright © 2012 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 transport demands at a given junction station fluctuate obviously in different time periods, which makes the rail departments unable to establish an even train operation schedule. This paper considers an optimization problem for train operations at the junction station in passenger dedicated lines. A satisfaction function of passengers is constructed by means of analyzing the satisfaction characteristics and correlative influencing factors. Through discussing the passengers' travel choice behavior, we formulate an optimization model based on maximum passenger satisfaction for the junction and then design a heuristic algorithm. Finally, a numerical example is provided to demonstrate the application of the method proposed in this paper.

1. Introduction

In passenger dedicated lines, passenger trains feature high-speed, high-density, and small train-unit, and the characteristics of passenger transport demands are similar to those of city bus passenger demands. Therefore, train operations in passenger dedicated lines need to be designed differently from the cases in other lines. In particular, train operations at junction stations must be schemed based on the changing of passenger demands. Many scholars had studied the train plan problem in passenger dedicated lines. Crisalli presented a system of within-day dynamic railway service choice and assignment models, which were used as a large decision support system for the operational planning of rail services [1]. Salido and Barber presented a set of heuristics for a constraint-based train scheduling tool to formulate train scheduling as constraint optimization problems [2]. Freling et al. introduced the problem of shunting passenger train units in a railway station. Shunting occurs whenever train units are temporarily not necessary to operate a given timetable [3]. Some other scholars noticed that passenger flows were much uneven in different time-periods in everyday, so they studied the dynamic feature of passenger flows. Niu presented a matching problem between

skip-stop operations and time-dependent demands in city headways, and formulated a nonlinear programming model which minimized the overall waiting time and in-vehicle crowding costs [4]. Jha et al. studied the alternative travel choices which were evaluated by a disutility function of perceived travel time and perceived schedule delay, and formulated a Bayesian updating model to optimize an alternative scheme [5]. He et al. presented a fuzzy dispatching model to assist the coordination among multiobjective decisions in railway dispatching plan [6]. Nie et al. considered the passenger choice behavior in rail, and proposed a calculation method of network impedance which could reduce the influence of different travel distance in passengers choice behavior [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 the passenger's total travel time loss is the two planning objectives of the model [8]. Shi et al. established a multiobjective optimum model of passenger train plans for dedicated passenger traffic lines by balancing benefits of both railway transportation corporations and passengers, and proposed a method to solve the optimization problem [9]. Ghoseiri et al. developed a multiobjective optimization model for the passenger train-scheduling problem on a railroad network which included single and multiple tracks [10]. In this study, lowering the fuel consumption cost was the measure of satisfaction of the railway company, and shortening the total passenger-time was being regarded as the passenger satisfaction criterion.

In previous studies, operation plans of passenger trains were mainly studied on optimizing transport organization in rail lines, including the train stop-schedule, service frequency, fleet size, and so forth. However, research on optimizing transport organization at a junction station is less concerned by other scholars. In general, the optimization objective of train operations for rail departments is to utilize trains efficiently and to lower travel cost for passengers. Therefore, the optimized train operations should be balanced between rail departments' operating cost and the travel cost of passengers. However, passenger demands are uneven in different time-periods, and train-set quantity is limited at junction stations; the train-set quantity probably cannot meet passenger demands at peak time-period. Passengers will be unsatisfied when their travel cost is increased by longer waiting time or raised ticket price. Thus, it is important to reasonably arrange the train's quantity and degrees in different time-periods. In this paper, we will focus on the matching of passenger train quantity with passenger demands at junction stations in different time-periods in passenger dedicated lines.

This paper is organized as follows: Passenger demands and travel choices at peak time-period are discussed in Section 2. An optimized model is built in Section 3, whose objective is to get maximum total passengers' satisfaction at the junction stations. In Section 4, a heuristic algorithm is designed to meet the changing of passenger demands and satisfy the constraint of train-set quantity. A numerical example is provided to illustrate the application of the model and algorithm in Section 5. The last section highlights the conclusion, and suggests future research directions.

2. Problem Statement

2.1. Passenger Demands and Travel Choices

As mentioned above, the passenger demands in passenger dedicated lines are heavily different at different time-periods, thus train operations show irregularly in every time-period at junction stations. Previous studies have discussed the departing interval of trains at

junction stations based on the condition that passenger transport demands and capacities are equal during the peak time-periods. In fact, the hypothesis is unreasonable when passenger transport demands are larger than capacities during the peak time-periods. Therefore, the railway departments can not arrange enough trains to meet the passenger demands in the peak time-periods.

Passengers mainly consider the degree and departing time of trains when they choose railway to travel. However, passenger demands are larger than transport ability during the peak time-periods. In this case, this paper considers that passengers probably choose the following suboptimal travel schemes for themselves: avoiding the peak time-periods and taking a train of the same degree to travel in another time-period; choosing a train of another degree to travel when its quantity is large enough in the same time-period.

2.2. Optimization Methods

The optimization objective for train operations at junction stations is to get maximum passenger satisfaction. Passenger satisfaction for train operations, presented in this paper, is an important indicator to evaluate the train operations. Here, the service time at junction stations is divided into m time-periods, according to which passenger demands and train-set quantity at junction stations are obtained, respectively. The time-periods in which passenger demands are larger than transport abilities are defined as peak time-periods. Finally, train operations in peak time-periods are organized according to the following two schemes.

2.2.1. Transferring Passenger Demands

The process that passengers choose the suboptimal schemes to travel is shown in Figure 1. The parameters u and u' represent some time-periods in the service time at the junctions, respectively; a and c indicate the aboard process that passengers take r and r' degree trains, respectively; b and d indicate the travelling process that passengers take r and r' degree trains, respectively; e represents the fare loss of passengers who intend to take r' degree train but are transferred to r degree train; f represents travelling time loss of the passenger who intends to take r degree train but is transferred to r' degree train; g indicates the waiting time cost of passengers who have to travel in time-period u' .

2.2.2. Adjusting Operation Section of Train-Set

Passenger demands generate at the junction stations, from which passenger trains are dispatched to different terminal stations j_1 , j_2 , and j_3 in passenger dedicated lines as shown in Figure 2.

As mentioned above, the transport capacity is limited at a junction station because the train-set quantity is affected by its operation mode. Stationary operation mode of train-set is used in most of the passenger dedicated lines presently; train-sets are operated on fixed railway sections as shown in Figure 3. s represents a train-set; t_f represents the departure interval of the same kind of train-sets at stations. In general, the value of t_f is larger than that of t_z , which represents the train service time at stations as shown in Figure 3(a). However, the value of t_f must be minimized to just meet train servicing time at stations; the value of t_f

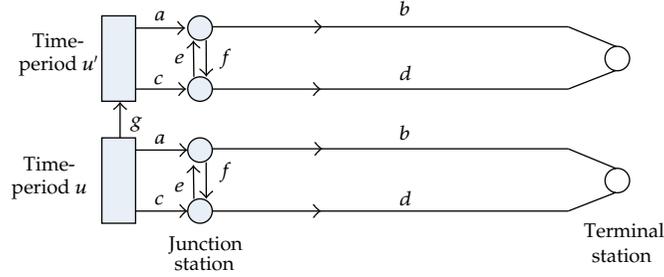


Figure 1: The process that passengers choose suboptimal schemes to travel.

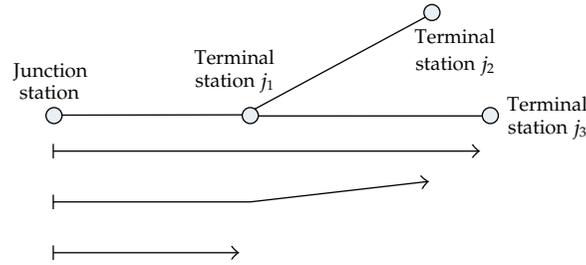


Figure 2: Through train plan from original station to different terminal stations.

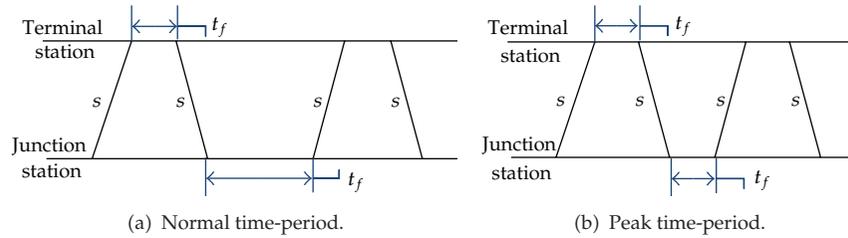


Figure 3: The operation mode of train-sets on fixed railway section.

is equal to that of t_z in peak time-period as shown in Figure 3(b). Thus, the measure ensures that the train-set operation will be optimized and the transport capacity will be raised.

Train-sets are not utilized completely in normal periods according to the above analysis. The paper introduces a method to adjust some train-sets from one rail section to another. For example, u is the peak time-period for rail section 1 but not for rail section 2. Moreover, there are superfluous train-sets in rail section 2, then we can adjust some from section 2 to section 1 in time-period u .

3. The Train Operation Model

3.1. Definitions and Notations

The following notations are used to describe the proposed model.

U is the set of time-periods, $U = \{1, 2, \dots, m\}$; $u, u' \in U$, $u \neq u'$.

R is the set of train degrees, $R = \{1, 2\}$; $r, r' \in R, r \neq r'$.

J is the set of terminal stations, $J = \{j_1, j_2, j_3\}$, $j \in J$.

$m_j^{u,r}$ is the demands of passengers who prepare to get r th degree trains to j th terminal station in u th time-period.

$e_j^{u,r}$ is the train-set quantity of r th degree trains which are dispatched to j th terminal station in u th time-period.

k_r is the maximum number of seats in r th degree trains.

$\delta_j^{u,u'}$ is the passenger satisfaction when passenger's departure time is changed from u th to u' th time-period.

$\delta_j^{r,r'}$ is the passenger satisfaction when passenger train is changed from r th to r' th degree.

$\xi_j^{u,r}$ is the weighted average satisfaction of total passengers who prepare to get to j th terminal station by r th degree trains in u th time-period.

Three intermediate variables are defined as follows.

$x_j^{u,r}$ is the passenger demands which can be contented in $m_j^{u,r}$.

$y_j^{u,r}$ is the passenger demands which are adjusted to travel in u' th time-period.

$z_j^{u,r'}$ is the passenger demands which are adjusted to travel by r' th degree trains.

The decision variable is defined as follows.

$n_j^{u,r}$ is the train operation quantity when rail department organizes r th degree trains to j th terminal station in u th time-period.

3.2. Passenger Satisfaction Function

3.2.1. Passenger's Sensitivity for Changing Their Travel Plan

For every passenger who prepares to travel by r th degree train in u th time-period, their satisfactions are different. In this paper, the satisfaction value is set to 1 when the passenger's travel plan is contented; otherwise, the value is smaller than 1. Passenger satisfactions are on account of passenger's sensitivity for the changing of their departure time or train degree. The passenger's sensitivity for departure time is a tolerable degree for waiting time when passengers have to change their departure time to travel. The passenger's sensitivity for train degree is a tolerable degree for ticket price rise when passengers have to change the train degree to travel.

The function $f_j(u, u')$ is defined to illustrate the passenger's sensitivity when their departure time is delayed. It is related with the waiting time and travel time as shown in formula (3.1). The formula reflects a ratio relation between the waiting time at junction stations and the travel time in lines. In this formula, numerator represents the passenger's waiting time at junction stations, and denominator shows the travel time on lines. The passenger's departure time is put off to the next time-period when $|u - u'| = 1$. Otherwise, $|u - u'| > 1$. Consider

$$f_j(u, u') = \frac{|u - u'| \cdot t_p}{t_j}, \quad (3.1)$$

where t_p is the average waiting time of every time-period. t_j represents the travel time from junction station to the j th terminal station.

The function $g_j(r, r')$ is defined to illustrate the passenger's sensitivity when their train degree is changed. It is related with the rangeability of ticket price, as shown in (3.2). The value of p_j^r is larger than that of $p_j^{r'}$ when the value of r is less than that of r' . The passenger's satisfaction does not decrease with the reducing of ticket price in this case but will decrease with the changing of their travel plan. Here, λ is introduced to present the descending of passenger satisfaction:

$$g_j(r, r') = \begin{cases} \lambda & r < r', \\ \frac{p_j^{r'} - p_j^r}{p_j^r} \cdot \lambda & r > r', \end{cases} \quad (3.2)$$

where, λ is the parameter when the train degree is changed, $0 \leq \lambda \leq 1$. p_j^r represents the ticket price of r th degree train from the junction station to j th terminal station.

3.2.2. Passenger Satisfaction for Changing Travel Plan

The value of $\delta_j^{u, u'}$ is correlative with the function $f_j(u, u')$. The larger the value of $f_j(u, u')$, the smaller the value of $\delta_j^{u, u'}$. The passenger satisfaction function for changing train departure time can be defined as (3.3). In the same way, $\delta_j^{r, r'}$ has similar character, and the passenger satisfaction function for changing train degree is defined as (3.4):

$$\delta_j^{u, u'} = \exp[-f_j(u, u')], \quad (3.3)$$

$$\delta_j^{r, r'} = \exp[-g_j(r, r')]. \quad (3.4)$$

3.2.3. Passenger Satisfaction Function

In this paper, passenger satisfaction is defined as formula (3.5) representing the weighted average satisfaction of total passengers who prepare to get to the j th terminal station by the r th degree train in the u th time-period:

$$\xi_j^{u, r} = \frac{x_j^{u, r} \cdot 1 + \sum_{u' \in U, u' \neq u} y_j^{u', r} \cdot \delta_j^{u, u'} + z_j^{u, r'} \cdot \delta_j^{r, r'}}{m_j^{u, r}}. \quad (3.5)$$

3.3. The Division Method of Time-Period

According to the above analysis, passenger satisfaction will decrease when passenger's waiting time is enlarged at junction station, as passenger's waiting time will increase with the prolonging of the time-period. Thus, time-periods are divided according to the minimum passenger satisfaction τ . The time-period division is unreasonable when the value of $\delta_j^{u, u'}$ is less than τ . This paper computes the value of t_p when the parameter $\delta_j^{u, u'}$ is equal to τ , and

uses the value of t_p as the dividing standard of time-period, as shown in formula (3.6). The number of time-periods is calculated by (3.7):

$$t_p = -\frac{t_j}{|u - u'|} \ln \tau, \quad (3.6)$$

$$m = \frac{l}{t_p}, \quad (3.7)$$

where m is the number of time-periods, and l denotes the length of service time in passenger dedicated lines.

3.4. Modeling

3.4.1. Objective Function

Here, $\xi_j^{u,r}$ represents the weighted average satisfaction of total passengers who prepare to get to j th terminal station by r th degree train in u th time-period, and the range of value $\xi_j^{u,r}$ is from 0 to 1. In (3.8), the objective is to get maximum total passenger satisfaction:

$$\max Q = \sum_{j \in J} \sum_{u \in U} \sum_{r \in R} \xi_j^{u,r}, \quad (3.8)$$

where, $\sum_{r \in R} \xi_j^{u,r}$ represents the satisfaction of passengers who get to j th terminal station at u th time-period. $\sum_{u \in U} \sum_{r \in R} \xi_j^{u,r}$ is the satisfaction of all passengers who get to j th station.

3.4.2. Constraints

The constraint of even passenger flow is shown in (3.9). The value of $m_j^{u,r}$ can be divided into $x_j^{u,r}$, $y_j^{u,r}$, and $z_j^{u,r}$ when the passenger demands cannot be contented completely in peak-periods. $\sum_{u' \in U, u' \neq u} y_j^{u',r}$ indicates the demand of passengers who prepare to travel in u th time-period and probably to be assigned to other time-periods. Consider

$$x_j^{u,r} + \sum_{u' \in U, u' \neq u} y_j^{u',r} + z_j^{u,r} = m_j^{u,r}. \quad (3.9)$$

The constraint of the train operation quantity balance is shown in (3.10). In this formula, $\varepsilon = \{x_j^{u,r}, \sum_{u' \in U} y_j^{u',r}, z_j^{u,r}\}$, "mod" is the symbol of modular division. The formula " $x_j^{u,r} \bmod k_r + \beta_\varepsilon \cdot 1$ " indicates that the train quantity should meet the passenger demands

$x_j^{u,r}$. Similarly, the formula “ $\sum_{u' \in U} y_j^{u',r} \bmod k_r + \beta_\varepsilon \cdot 1$ ” and “ $z_j^{u,r'} \bmod k_r + \beta_\varepsilon \cdot 1$ ” represent the train quantity meeting the demands $\sum_{u' \in U, u' \neq u} y_j^{u',r}$ and $z_j^{u,r'}$, respectively:

$$n_j^{u,r} = x_j^{u,r} \bmod k_r + \sum_{u' \in U} y_j^{u',r} \bmod k_r + z_j^{u,r'} \bmod k_r + \sum_{\varepsilon} \beta_\varepsilon \cdot 1, \quad (3.10)$$

$$\beta_\varepsilon = \begin{cases} 1 & \varepsilon \bmod k_r \neq 0, \\ 0 & \varepsilon \bmod k_r = 0. \end{cases} \quad (3.11)$$

The constraint of the train-set quantity is shown in (3.12), which represents that train quantity $n_j^{u,r}$ is restricted by train-set quantity $e_j^{u,r}$:

$$n_j^{u,r} \leq e_j^{u,r}. \quad (3.12)$$

The constraint of minimum passenger satisfaction is shown in (3.13), in which $\delta_j^{u,u'}$ and $\delta_j^{r,r'}$ should be larger than the empirical value of the passenger's toleration for changing travel plan. Consider

$$\begin{aligned} \delta_j^{u,u'} &\geq \tau, \\ \delta_j^{r,r'} &\geq \tau. \end{aligned} \quad (3.13)$$

The nonnegative and integer constraint is shown in:

$$n_j^{u,r} \geq 0, \quad x_j^{u,r} \geq 0, \quad y_j^{u',r} \geq 0, \quad z_j^{u,r'} \geq 0 \quad (3.14)$$

and are integer.

4. Algorithm Design

This paper designs a heuristic algorithm of train operation based on maximum passenger satisfaction. The algorithm process is shown as follows.

Step 1 (initialization). Firstly, the smaller value between train-set and train demand quantity is assigned to the train operation quantity, namely $n_j^{u,r} = \min\{e_j^{u,r}, a_j^{u,r}\}$. Secondly, the value of demands $m_j^{u,r}$ is assigned to intermediate variable $x_j^{u,r}$, and 0 is assigned to intermediate variable $y_j^{u',r}$ and $z_j^{u,r'}$, respectively. Thirdly, define the counter b and feasible scheme p . Finally, the value of b and Q are set to 0.

Step 2 (examining the balance constraint of train-set capacity and demand). If the train demand quantity $a_j^{u,r}$ is less than train operation quantity $n_j^{u,r}$ in some time-periods, namely $m_j^{u,r} \bmod k_r + \beta_\varepsilon \cdot 1 < n_j^{u,r}$, go to Step 3. Otherwise, go to Step 4.

Table 1: The train ticket price.

r, j	1, 1	1, 2	1, 3	2, 1	2, 2	2, 3
P_j^r	80	120	135	50	80	100

Table 2: Passenger demands (unit: person times).

u, r	1, 1	1, 2	2, 1	2, 2	3, 1	3, 2	4, 1	4, 2
$m_{j_1}^{u,r}$	1110	3600	5500	12200	11100	6100	2780	1220
$m_{j_2}^{u,r}$	8300	9100	13000	9150	10000	3650	2750	2440
$m_{j_3}^{u,r}$	5500	4880	11100	9150	5500	7300	1660	1800

Table 3: The value of $e_j^{u,r}$ and $a_j^{u,r}$ (unit: train).

u, r	1, 1	1, 2	2, 1	2, 2	3, 1	3, 2	4, 1	4, 2
$e_{j_1}^{u,r}, a_{j_1}^{u,r}$	10, 2	8, 6	5, 10	10, 20	15, 20	15, 10	7, 5	5, 2
$e_{j_2}^{u,r}, a_{j_2}^{u,r}$	14, 15	13, 15	20, 25	25, 15	20, 18	10, 6	5, 5	2, 4
$e_{j_3}^{u,r}, a_{j_3}^{u,r}$	12, 10	10, 8	15, 20	20, 15	7, 10	10, 12	8, 3	8, 3

Step 3 (passenger satisfaction examination). Firstly, take the value of q meeting $q = \min\{\delta_j^{u,u}, \delta_j^{r,r'}\}$. Secondly, passenger satisfaction is calculated when the value of q is larger than that of τ , and output this scheme p . Otherwise, all passenger demands transformed from other time-periods are adjusted to prior time-period $u-1$, and equal passenger demands generated at time-period $u-1$ are adjusted to u th time-period. Then go to Step 4. Secondly, the counter p is refreshed with the equation of $p = p + 1$.

Step 4 (adjusting scheme). The value of $b_j^{u,r}$ is assigned to the value of $a_j^{u,r} - n_j^{u,r}$ when the value of $a_j^{u,r}$ is larger than $n_j^{u,r}$. Then, the value of $b_j^{u,r}$ is assigned equally to $y_j^{u,r'}$ and $z_j^{u,r'}$, according to formula $y_j^{u,r'} = z_j^{u,r'} = 0.5k_r b_j^{u,r}$. Finally, the value of train-set quantity $n_j^{u,r}$ is refreshed by formula $n_j^{u,r} = (x_j^{u,r} + y_j^{u-1,r} + z_j^{u,r'}) \bmod k_r + \beta_e \cdot 1$, and go to Step 2.

5. Numerical Example

In some passenger dedicated lines, passenger trains are only operated from the junction station to the terminal stations j_1 , j_2 , and j_3 . The travel time from the junction station to terminal stations j_1 , j_2 , and j_3 is 2, 3, and 5 hours, respectively, namely $t_1 = 2$, $t_2 = 3$, and $t_3 = 5$. There are two degree trains, $r = 1$, $r' = 2$. The trains' ticket prices are shown in Table 1. The service time of every day is 14 hours from 6:00 to 20:00. The length of the time-period t_p and the number of time-periods m are computed according to (3.6) and (3.7). The calculation results: the value of t_p is 3.3 hours, and the number of time-period m is 4. The service time can be divided into four time-periods, from which the passenger demands collected are shown in Table 2.

The passenger demands can be transformed to train demands according to (3.10), and the constraint of train-set is given in the numerical example as shown in Table 3, where the notation $e_j^{u,r}$ and $a_j^{u,r}$ represent maximum train-set quantity and train demands. Then the

Table 4: Train operation quantity in every time-period $n_j^{u,r}$ (unit: train).

u, r	1,1	1,2	2,1	2,2	3,1	3,2	4,1	4,2
j_1	2	6	5	10	20	15	7	6
j_2	13	11	20	23	20	6	5	2
j_3	10	8	15	20	7	10	6	5

Table 5: Adjustment the scheme of passenger trains.

Terminal station	Adjustment program [(u, r) \rightarrow (u', r) or (u, r'): the adjusted train quantity]
j_1	(2,1) \rightarrow (3,1): 4, (2,2) \rightarrow (3,2): 6, (3,1) \rightarrow (4,1): 5
j_2	(1,1) \rightarrow (1,2): 2, (2,1) \rightarrow (3,1): 3, (1,1) \rightarrow (3,1): 1, (4,2) \rightarrow (4,1): 1
j_3	(2,1) \rightarrow (3,1): 5, (3,1) \rightarrow (4,1): 2, (3,2) \rightarrow (4,2): 2

Table 6: Passenger satisfactions.

u, r	1,1	1,2	2,1	2,2	3,1	3,2	4,1	4,2
j_1	1	1	0.53	0.7	0.85	0.93	1	1
j_2	0.87	0.73	0.8	0.93	0.89	0.8	1	0.5
j_3	1	1	0.75	1	0.7	0.83	1	1

datum, in which $e_j^{u,r}$ is less than $a_j^{u,r}$, is adjusted to other time-periods or other degree trains according to the above heuristic algorithm.

This paper optimizes the passenger operation at the junction station according to the above model and algorithm in Table 4. The adjustment result of passenger demands is shown in Table 5, and that of the passenger satisfaction is in Table 6.

6. Conclusion

In this paper, an optimization model based on maximum passenger satisfaction for the junction station has been given. A heuristic algorithm is proposed to solve it. According to the scheme results, all passenger satisfaction is calculated. Average satisfaction of passengers who prepare to get to j_1 th, j_2 th, and j_3 th terminal stations are 0.87, 0.81, and 0.91, respectively. Minimum satisfaction of passengers who prepare to get to j_1 th, j_2 th, and j_3 th terminal station are 0.42, 0.45, and 0.4, respectively. The result shows that the method proposed in this paper can effectively solve the problem, and is suitable for formulating passenger train operation in passenger dedicated lines. Furthermore, it is an important topic for further research to consider the train operation based on collaborative optimization among several junction stations in passenger dedicated lines.

Acknowledgments

The work described in the paper was supported by National Nature Science Foundation of China under Grant no. 50968009 and no. 71261014, and the Research Fund for the Doctoral Program of Higher Education under Grant no. 20096204110003.

References

- [1] U. Crisalli, "User's behaviour simulation of intercity rail service choices," *Simulation Practice and Theory*, vol. 7, no. 3, pp. 233–249, 1999.
- [2] M. A. Salido and F. Barber, "Mathematical solutions for solving periodic railway transportation," *Mathematical Problems in Engineering*, vol. 2009, Article ID 728916, 19 pages, 2009.
- [3] R. Freling, R. M. Lentink, L. G. Kroon, and D. Huisman, "Shunting of passenger train units in a railway station," *Transportation Science*, vol. 39, no. 2, pp. 261–272, 2005.
- [4] H. M. Niu, "Determination of the skip-station scheduling for a congested transit line by bilevel genetic algorithm," *International Journal of Computational Intelligence Systems*, vol. 6, no. 4, pp. 1158–1167, 2011.
- [5] M. Jha, S. Madanat, and S. Peeta, "Perception updating and day-to-day travel choice dynamics in traffic networks with information provision," *Transportation Research Part C*, vol. 6, no. 3, pp. 189–212, 1998.
- [6] S. He, R. Song, and S. S. Chaudhry, "Fuzzy dispatching model and genetic algorithms for railyards operations," *European Journal of Operational Research*, vol. 124, no. 2, pp. 307–331, 2000.
- [7] L. Nie, X. F. Hu, L. Tong, and H. L. Fu, "Research of passenger flow assignment based on passenger train plan," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 3, pp. 87–92, 2011.
- [8] 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 Part B*, vol. 34, no. 2, pp. 91–106, 2000.
- [9] F. Shi, L. B. Deng, X. H. Li, and Q. G. Fang, "Research on passenger train plans for dedicated passenger traffic lines," *Journal of the China Railway Society*, vol. 26, no. 2, pp. 16–20, 2004.
- [10] K. Ghoseiri, F. Szidarovszky, and M. J. Asgharpour, "A multi-objective train scheduling model and solution," *Transportation Research Part B*, vol. 38, no. 10, pp. 927–952, 2004.

Research Article

Stochastic User Equilibrium Assignment in Schedule-Based Transit Networks with Capacity Constraints

Wangtu Xu,¹ Lixin Miao,² and Wei-Hua Lin³

¹ School of Architecture and Civil Engineering, Xiamen University, Xiamen 361005, China

² Research Center for Modern Logistics, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

³ Department of Systems and Industrial Engineering, The University of Arizona, Tucson, AZ 85721, USA

Correspondence should be addressed to Wangtu Xu, ato1981@xmu.edu.cn

Received 9 June 2012; Revised 23 August 2012; Accepted 26 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Wangtu Xu 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 stochastic user equilibrium (SUE) assignment model for a schedule-based transit network with capacity constraint. We consider a situation in which passengers do not have the full knowledge about the condition of the network and select paths that minimize a generalized cost function encompassing five components: (1) ride time, which is composed of in-vehicle and waiting times, (2) overload delay, (3) fare, (4) transfer constraints, and (5) departure time difference. We split passenger demands among connections which are the space-time paths between OD pairs of the network. All transit vehicles have a fixed capacity and operate according to some preset timetables. When the capacity constraint of the transit line segment is reached, we show that the Lagrange multipliers of the mathematical programming problem are equivalent to the equilibrium passenger overload delay in the congested transit network. The proposed model can simultaneously predict how passengers choose their transit vehicles to minimize their travel costs and estimate the associated costs in a schedule-based congested transit network. A numerical example is used to illustrate the performance of the proposed model.

1. Introduction

Transit assignment is an approach used for predicting the way in which passengers choose routes traveling from origins to destinations. Much progress has been made in the past three decades [1], and the assignment model can be broadly divided into three types: transport system-based, frequency-based, and schedule-based.

In a transport system-based network, the all-or-nothing assignment method in which passengers choose the quickest route without considering headways of line routes as well as timetables is adopted. The result provides an overview of the structure of travel demand

for long-distance planning purposes. Generally, the transport system-based assignment procedure does not require any line frequencies or timetables as input data. The early transit assignment approaches such as Dial's algorithm [2, 3] and the method by Fearnside and Draper [4] are also based on the transport system in a way similar to road traffic assignment.

In a frequency-based network, each transit line is assumed to operate on a constant headway. The assignment procedure encompasses three steps: route search, route choice, and demand split. The first step searches for possible paths between all origin-destination (OD) pairs. The second step compares the individual routes and eliminates the unreasonable routes. Then the final step evaluates the remaining routes and assigns the trips of an OD matrix to these routes.

During the route choice step, for at least some OD pairs, there are sections in a path with more than one parallel service offered and passengers can choose the one they perceive as the best, which leads to the common lines problem [5], often regarded as the most complex problem for transit assignment. De Cea et al. [6] proposed an alternative method of generating minimum cost routes, as well as the partial paths from different lines using a common route section with a nonlinear programming method. Following the ideas of Chriqui and Robillard [5], Spiess [7], Spiess and Florian [8] introduced a strategy for choosing an attractive route set of lines at boarding stop points. This idea was further extended by Wu et al. [9] who proposed the strategy-based asymmetric transit link cost function and the hyperpath concept. These models assumed no capacity limit for links of a network. Gendreau [10] was the first to formulate a general transit assignment with the capacity constraint, and following by Lam et al. [11], Cominetti and Correa [12] and Kurauchi et al. [13]. Lam et al. advanced a stochastic user equilibrium assignment model for congested transit networks with a solution algorithm that can simultaneously predict how passengers choose their optimal routes and estimate the total passenger travel cost [11]. Cominetti and Correa proposed a model based on the common lines paradigm, which was applied to general networks using a dynamic programming approach, and congestion was treated by means of a simplified bulk queue model [12]. Kurauchi et al. proposed a model in which passengers unable to board due to the capacity constraint were then routed through spill-links [13]. These algorithms considered the congestion situation by introducing a volume-dependent link cost function with the capacity constraint. Consequently the resulting equilibrium models could be solved by standard algorithms for convex minimization. Another important topic on frequency-based transit assignment is the common line problem based on the hyperpath approach. Nguyen et al. [14] investigated the application of a nested logit model to trip assignment on urban transit networks where every set of competitive transit lines is described by a hyperpath. Schmöcker [15] also employed the hyperpath concept to the transit assignment problem with the capacity constraint.

In general, the frequency-based transit assignment algorithm assumes that the passenger demand is constant within the specified time period of interest. The transfer time is not explicitly calculated but to be estimated based on the headway of the transit vehicle. This means that the impact of timetable is not considered, and the waiting time is usually assumed to be equal to the half of the headway.

On a schedule-based transit network, the assignment considers the exact timetable and therefore the procedure needs to model the spatial and temporal structure of travel demand. The resulting assignment would show explicitly the exact number of passengers apportioned to each scheduled vehicle. Recently, this method becomes more and more popular, Florian [16] firstly proposed a deterministic schedule-based transit assignment method and applied it to the EMME/2 software package, in which the weight factors and non-time-based cost

elements in determining the optimal path were used to evaluate the feasibility of a path and its attractiveness, and the shortest path algorithm was employed to assign trips. Tong and Wong [17, 18] formulated a dynamic transit assignment model. In their model, passengers were assumed to travel on a path with minimum generalized costs. These algorithms could be applied over a period in which both passenger demands and vehicle headways are varying. Friedrich and Wekeck [19] constructed a transit path choice method using the branch and bound technique, which reduced further the computation time. Friedrich [20] made an extension of their algorithm from a single-day to a multiday situation, which allowed considering changes in supply and demand within the course of a multiday time period. Nielsen [21] developed a stochastic schedule-based transit assignment model considering the utility of different passengers and optimized the stochastic assignment model based on the method of successive averages (MSA) [22]. Nuzzolo also developed algorithms for the transit assignment problem [23, 24]. Xu et al. proposed the K -shortest path searching algorithm in a schedule-based transit network [25]. This algorithm could be used in the flow assignment when the time-space path is taking part in the flow split between the OD pair. In all those algorithms developed, the attractive connection in a schedule-based network is not considered. Besides, the stochastic path choice behavior in the congested situation has not been studied.

Previous studies involved in flow assignment methods in the schedule-based transit network are extremely limited, let alone the consideration of capacity constraint in a stochastic user equilibrium (SUE) transit assignment model. In this paper, a schedule-based SUE transit assignment algorithm for a similar common line problem is presented. We consider the schedule-based transit network described in Friedrich and Wekeck [19]. In our work, however, we assume that passengers do not necessarily have full knowledge of the schedule of transit service. A stochastic user equilibrium assignment method with the congested situation is proposed, which is an extension of the work by Lam et al. [11]. The latter considers the problem in the context of a frequency-based transit network. The purpose of this paper is to formulate a model to determine exactly the load of vehicle on a transit line at a given time period. Moreover, we examine whether the passenger volume on a transit line exceeds the designed capacity.

This paper is organized as follows. In the next section, some useful concepts for a transit network are briefly reviewed. Notations and basic assumptions of the mathematical model are given. In Section 3, the attractive connection set is defined. In Section 4, a generalized travel cost function is formulated to choose the best routes between OD pairs on the schedule-based transit network firstly. Then a SUE assignment model is proposed, as well as its solution algorithm. The numerical example of this model is presented in Section 5 to illustrate the validity of the algorithm. Conclusions are given in Section 6 as well as the direction for future research.

2. Concepts, Notations, and Basic Assumptions

2.1. Concepts

We provide here an overview of terms used in this paper such as line, line section, and line route before embarking on a discussion of the common lines problem and SUE transit assignment. We adopt the definitions of these terms from the previous work: definitions of the transit line, transit arc, and line segment from Lam et al. [11], definitions of route segment, connection and connection segment from Friedrich and Wekeck [19].

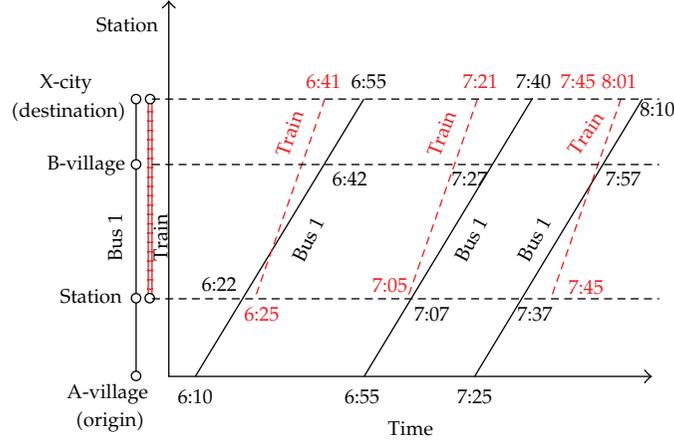


Figure 1: An example schedule-based transit network (Friedrich and Wekeck [19]).

Table 1: Connections of example schedule-based transit network.

Connection ID	Passing stations	Departure time	Arrival time	In-vehicle time	Transfer time	Total travel time	Number of transfers
c_1	A-S-B-X	6:10	6:55	45 min	0 min	45 min	0
c_2	A-S-B-X	6:55	7:40	45 min	0 min	45 min	0
c_3	A-S-B-X	7:25	8:10	45 min	0 min	45 min	0
c_4	A-S-X	6:10	6:41	28 min	3 min	31 min	1
c_5	A-S-X	7:25	8:01	28 min	8 min	36 min	1

A connection is a line that a passenger chooses to travel from his/her origin node to the destination node. In the schedule-based transit network, each connection is composed of origin node, destination node, walking link, transfer nodes, departure time, arrival time, in-vehicle time, transfer time, number of transfers, and the total travel time.

A connection segment is a portion of a connection which describes a part of a journey and is also endowed with a departure and arrival time, and which is the building block of a connection. In this paper, the connection segments using access and egress walk links would not be considered.

The example network (shown in Figure 1) used by Friedrich and Wekeck [19] is adopted here to explain the definition of line (segment), line route (segment) and connection (segment). In this example, the given network consists of a bus line and a train line, passengers traveling from origin A-Village to destination X-City may choose between direct bus connections and faster bus-train connections.

Since a transit line is characterized by an initial stop node, a terminal stop node, length and running time, the connection segments can be calculated from a set of line sections. For the example network given in Figure 1, bus line 1 from node A-Village to station has three connection segments, that is, boarding and departing at node A-Village at 6:10, arriving at station at 6:22; or departing at 6:55, arriving at 7:07; or departing at 7:25 and arriving at 7:37. A given transit line with n stations would include $n(n-1)/2$ line sections and consequently include a total of $k \times n(n-1)/2$ connection segments (where k is the total number of vehicle runs). The connections of the example transit network of Figure 1 are shown in Table 1.

From the above example, we can conclude that in a schedule-based transit network a connection is a route path plus a series of time strategies, including the departure time, arrival time and transfer time. When an incoming bus is operating at its capacity level, passengers may choose not to board but opt to wait for the next one with successor connections. In this situation, we can call this attractive connection problem for a schedule-based transit network with capacity constraints. Namely, passengers choose their journey strategy at a station from an attractive connection set.

2.2. Notations

W : Set of OD pairs

w : An element of set W

g_w : Passenger demand between OD pair w

C_w : Set of attractive connections associated with OD pair w

c : Index of connection

S : Set of connection segments of the attractive connections

s : Index of connection segment

v_c : Passenger flow on connection c

v_s : Passenger flow on connection segment s

d_c : Passenger overload delay, that is, the time that passengers spend on waiting for vehicle of another connection segment when they cannot board the first coming vehicle of the first connection segment because of insufficient vehicle capacity

d_s : Passenger overload delay on connection segment s

tt_c : Travel time on connection c

tt_s : Travel time on connection segment s

dt_c : Departure time of connection c

dt_s : Departure time of connection segment s

dt_E : Expected departure time of a trip

at_c : Arrival time of connection c

at_s : Arrival time of connection segment s

gc_c : Generalized cost of connection c

gc_s : Generalized cost of connection segment s

I : Analysis time span

t_c^w : Waiting time on connection c

t_s^w : Waiting time on connection segment s

t_c^v : In-vehicle time on connection c

t_s^v : In-vehicle time on connection segment s

t_c^d : Riding time on connection c

t_s^d : Riding time on connection segment s

- t_c^n : Number of transfers of connection c
- t_s^n : Number of transfers of connection segment s
- $T_B^{-(+)}$: The minimum (maximal) transfer wait time T_B^-
- d_c : Passenger overload delay on connection c
- d_s : Passenger overload delay on connection segment s
- Y_c : Monetary cost of connection c
- Y_s : Monetary cost of connection segment s
- $\xi_{s(c)}^I$: A nonnegative function of $\text{dist}(dt_{s(c)}, dt_E)$, which represents the functions of difference between the real departure time $dt_{s/c}$ and the expected departure time dt_E
- cap_{v-s}^l : The vehicle capacity of transit line l
- δ_c^w : 0-1 variable, if connection c connects between OD pair w , it equals to 1 and otherwise 0.
- $\delta_{s,c}^w$: 0-1 variable, it equals to 1 if connection c associating OD pair w consists of connection segment s , and 0 otherwise.

2.3. Basic Assumptions

Assume that (a) the transit service considered has a schedule but passengers do not know exactly the schedule. The service behaves as if it was frequency based, (b) the OD demand is fixed, at any given time interval (e.g., the peak hour), and (c) all vehicles strictly operate under the sequence defined by the timetable without overtaking each other. (d) Other assumptions are drawn as Lam et al. [11].

3. The Attractive Connection Set

De Cea and Fernandez [26] indicated that in a congested transit network, there exists more than one type of route segments between a given pair of nodes representing the set of desirable lines. However, in a schedule-based network, there would also be more than one type of connection segments which would result in an attractive connection set as described above. Different from the determination of different classes of route types of uncongested network by solving the hyperbolic common lines problem [26], an important process needs to be emphasized to determine the most attractive connection set. This will be discussed as follows.

According to the timetable, the frequency of a transit vehicle is fixed. Assume that the preceding vehicle would not be overrun, the set of attractive connections based on some specific rules could be determined. The algorithm used by Friedrich and Webeck [19] is employed to build a connection tree. Every connection segment s is described by departure time dt_s and arrival time at_s , travel time tt_s , cost cost_s , and number of transfer t_s^n . If connection c is made up of n connection segments, the number of its transfers is $n - 1$. As shown in Figure 2 [19], let s be the current processed connection segment from station B to station C , let c^* be the new connection between the original station A and station C formed by adding s to some connection c arriving at B , finally let C_{S-C} be the set of all known connections to

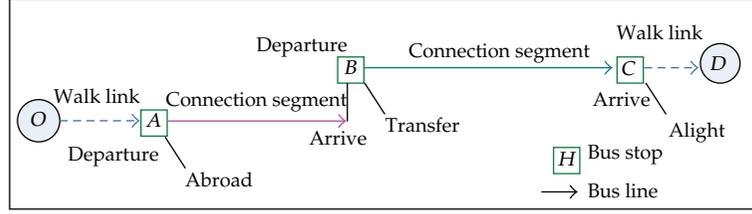


Figure 2: Structure of connection.

stop C. Connection segment s is inserted into the tree as a successor of c , if and only if the following conditions hold.

- (1) Temporal suitability: The connection segment s departs from node B only after the arrival of connection c plus a minimum transfer wait time T_B^- , and before the maximum transfer wait time T_B^+ has elapsed, namely.

$$dt_s - at_c \in [T_B^-, T_B^+]. \quad (3.1)$$

- (2) Dominance: there is no known connection $c' \in C_{S-C}$, such that

$$dt_{c'} \geq dt_{c^*}, \quad at_{c'} \leq at_{c^*}, \quad gc_{c'} \leq gc_{c^*}, \quad t_{c'}^n \leq t_{c^*}^n. \quad (3.2)$$

- (3) Tolerance constraints: none of the following rules are violated:

$$gc_{c^*} \leq b_1 \times \min\{gc_{c'}, \forall c' \in C_{S-C}\} + b_2, \quad t_{c^*}^n \leq d_1 \times \min\{t_{c'}^n, \forall c' \in C_{S-C}\} + d_2 \quad (3.3)$$

$$tt_{c^*} \leq \varphi_1 \times \min\{tt_{c'}, \forall c' \in C_{S-C}\} + \varphi_2, \quad t_{c^*}^n \leq N^+,$$

where b_i, d_i, φ_i are user-defined global tolerance parameters, and N^+ is the user-defined bound for the number of transfers within a connection.

Using this approach which is the same as the branch and bound algorithm, we can determine all connections between any two nodes of the network. For a given OD pair (or any two nodes), we can then determine the attractive connection set according to some attributes of connection, for instance, the number of transfers or the departure time. As an example, we can set the attractive connection set C_{S-C} from node A to node C in Figure 2 as all connections of which the departure times are earlier than 8:00, or, all the connections of which the number of transfers do not exceed 3.

4. SUE Assignment Model for Schedule-Based Transit Network with Capacity Constraints

4.1. Generalized Travel Time Cost Function of a Connection

Waiting at a transit station can be described as a process like this: at a given time interval I , consider a passenger heading towards the destination, whose original node is r and terminal

node t . To exit from r he/she can use the connection segment s to reach the next station j . The decision faced at station r is determined via the generalized cost $g_{c_s}(v_s)$ corresponding to the services operating on the connection segment s such that

$$g_{c_s}(v_s) = \alpha t_s^W + \beta t_s^V + \varphi t_s^n + \gamma d_s + \rho \xi_s^I + \varpi Y_s = dt_s^d + \varphi t_s^n + \gamma d_s + \rho \xi_s^I + \varpi Y_s, \quad (4.1)$$

where, parameters $\alpha = \beta = d$ and φ are user-defined factors. γ, ρ, ϖ are the conversion factors of time value.

The cost to reach at the destination using connection c can be determined as follows:

$$g_{c_c} = \begin{cases} \infty & \text{if } dt_c, at_c \notin I \\ \sum_{\forall s \in S} \delta_{s,c}^w g_{c_s} & \text{otherwise.} \end{cases} \quad \forall c \in C_w \quad (4.2)$$

4.2. Flow Conservation in a Congested Network

Passenger flows on connections which satisfy the following constraints. For each OD pair w , its trip demand g_w can be split into all possible attractive connections as

$$g_w = \sum_{c \in C_w} v_c. \quad (4.3)$$

Each connection segment s should satisfy the flow conservation of each specific connection c , that is

$$v_s = \sum_{w \in W} \sum_{c \in C_w} \delta_{s,c}^w v_c. \quad (4.4)$$

Furthermore, connection flow should satisfy the capacity constraint so that there would not be an overload in the transit vehicle to which connection segment s belongs

$$v_s \leq \text{cap}_{v-s}^I. \quad (4.5)$$

4.3. SUE Assignment Model Formulation

Definition 4.1. A SUE is achieved in a schedule-based transit network with capacity constraints when the allocation of passengers between alternative connections conforms to the following logit model:

$$-\theta(g_{c_c} - g_{c_{c'}}) = \ln\left(\frac{v_c}{v_{c'}}\right) \quad \forall c \in C_w, \quad (4.6)$$

where c and c' are the alternative connections associated with the same OD pair w , and $\theta > 0$ is a given parameter used to measure the degree of passengers' knowledge about the travel cost on a specific connection. In general, the corresponding θ value for schedule-based network

would be smaller than the transportation system based on frequency-based transit system. As $\theta \rightarrow \infty$, the result of SUE approximates that of user equilibrium (UE).

Based on (4.1) and (4.6), we have

$$\ln\left(\frac{v_c}{v_{c'}}\right) = -\theta(g_{c_c} - g_{c_{c'}}) = -\theta\left[d(t_c^d - t_{c'}^d) + \varphi(t_c^n - t_{c'}^n) + \gamma(d_c - d_{c'}) + \rho(\xi_c^I - \xi_{c'}^I) + \varpi(Y_c - Y_{c'})\right]. \quad (4.7)$$

As the total demand increases, the proportionate distribution of passenger flow between the two connections remains the same until one or more segments on either connection are overloaded.

We formulate the SUE assignment problem as follows:

$$(NP1) \min z = \sum_{s \in S} \int_0^{v_s} g_{c_s}(x) dx + \frac{1}{\theta} \sum_{w \in W} \sum_{c \in C_w} v_c \ln v_c, \quad (4.8a)$$

S.t.

$$g_w = \sum_{c \in C_w} v_c, \quad (4.8b)$$

$$v_s = \sum_{w \in W} \sum_{c \in C_w} \delta_{s,c}^w v_c, \quad (4.8c)$$

$$v_s \leq \text{cap}_{v-s}^l, \quad (4.8d)$$

$$v_c \geq 0, \quad w \in W, \quad c \in C_w. \quad (4.8e)$$

The Lagrangian function for NP1 can be formulated by

$$L = \sum_{s \in S} \int_0^{v_s} g_{c_s}(x) dx + \frac{1}{\theta} \sum_{w \in W} \sum_{c \in C_w} v_c \ln v_c + \chi_w \left(g_w - \sum_{c \in C_w} v_c \right) + \varepsilon_s (v_s - \text{cap}_{v-s}^l) - \mu_c^w v_c. \quad (4.9)$$

The Kuhn-Tucker conditions for problem NP1 can be formulated as follows:

$$\frac{1}{\theta} (\ln v_c + 1) + \sum_{s \in S} \delta_{s,c}^w \cdot g_{c_s}(v_s) - \chi_w + \varepsilon_c - \mu_c^w = 0, \quad (4.10)$$

$$-\mu_c^w v_c = 0, \quad (4.11)$$

where χ_w , ε_c and μ_c^w are the corresponding Lagrangian multipliers to (4.8b)–(4.8d).

We have that if $v_c > 0$ then $\mu_c^w = 0$; $g c_c = \sum_{s \in S} \delta_{s,c}^w \cdot g c_s(v_s)$ for all $c \in C_w$, so that we can easily formulate the following logit model for the connection flow split between OD pair w :

$$v_c = g^w \frac{\exp(-\theta \cdot g c_c + \varepsilon_c)}{\sum_{k \in C_w} \exp(-\theta \cdot g c_k + \varepsilon_k)} \quad (4.12)$$

For each connection segment, $\varepsilon_s = -\theta \gamma d_s$ is a condition for SUE transit assignment with bottlenecks. If θ is very large, the second term of the objective function of problem NP1 will become insignificant and hence this is an approximation to the UE problem.

4.4. Solution Algorithm for SUE Assignment Problem

There are several solution algorithms for the standard SUE assignment problem, such as the method of successive averages [27], the partial convex combination method [28], and the iterative balancing and convex combination method [29]. The SUE assignment problem with capacity constraints like NP1, however, cannot apply these approaches directly. Bell proposed an advanced method of successive average to solve a SUE road traffic assignment problem [30], which was adopted by Lam et al. [11] to solve the transit assignment problem like NP1 with bottlenecks. Based on the solution method of Lam et al. [11], we designed a method to solve the SUE assignment problem for schedule-based transit network.

Rewrite (4.10) as

$$v_c = \exp\left(-\theta\left(g c_c + \chi_w - \varepsilon_c + \frac{1}{\theta}\right)\right) = \exp(-\theta \cdot g c_c) \prod_{s \in C} E_s M_w, \quad (4.13)$$

where connection c connects OD pair w , $M_w = \exp(\chi_w + 1/\theta)$, while factor $E_s = \exp(-\varepsilon_s)$, a simple procedure is proposed to solve the SUE transit assignment problem NP1 with given OD flows at time interval I .

Step 1. $n = 1$. Calculate an attractive connection set, and corresponding $s \in S$ for all OD pairs $w \in W$ according to the algorithm mentioned in the previous Section. Set $E_s^{(n)} = 1$ for each $s \in S$, and $M_w^{(n)} = 1$ for $w \in W$.

Step 2. $n = 2$. If the convergence condition is satisfied, then stop; otherwise, for each $s \in S$, calculate

$$\begin{aligned} v_c(E^{(n)}, M^{(n)}) &= \exp(-\theta \cdot g c_c) \prod_{s \in C} E_s^{(n)} M_w^{(n)}, \\ \beta_s^{(n)} &= \frac{\text{cap}_{v-s}^l}{\sum_{w \in W} \sum_{c \in C_w} \delta_{s,c}^w v_c(E_s^{(n)}, M_w^{(n)})}, \\ E_s^{(n+1)} &= \min\left[1, \beta_s^{(n)} E_s^{(n)}\right]. \end{aligned} \quad (4.14)$$

For each $w \in W$, set

$$\beta_w^{(n)} = \frac{g_w}{\sum_{c \in C_w} \delta_c^w v_c(E^{(n)}, M^{(n)})}, \quad M_w^{(n+1)} = \beta_w^{(n)} M_w^{(n)}. \quad (4.15)$$

Step 3. $n = n + 1$, back to Step 2.

Step 4. For each $c \in C_w$, calculate

$$v_c^{(n)} = v(E^{(n)}, M^{(n)}). \quad (4.16)$$

For each $s \in S$, calculate $v_s^{(n)} = \sum_{w \in W} \sum_{c \in C_w} \delta_{s,c}^w v_c^{(n)}$, $d_s = -(\ln E_s) / \gamma \theta$.

For each $c \in C_w$, calculate $d_c = \sum_{s \in S} -(\ln E_s) / \gamma \theta$.

Step 5. For all OD pairs, if the gap function $G(n) = \sum_{w \in W} \sum_{c \in C_w} \delta_{s,c}^w |v_c^{(n+1)} - v_c^{(n)}| / \sum_{w \in W} g_w < \zeta$, then stop. (where ζ is an user-defined parameter).

5. Numerical Example

We test the NP1 problem with the example network shown in Figure 1. Let the capacity of bus line 1 be 200 persons/vehicle and the capacity of the train line 600 persons/train. Let fare of Bus 1 be 1.00 Rmb, and fare of Train be 3.00 Rmb. The OD demand between A-Village and X-City is 450 persons. The parameters for calculating the generalized cost of connections: $d = 1.0$; $\varphi = 0.5$; $\rho = 1.0$; $\varpi = 0.1$; $\xi_c^l = (dt_c - dt_E)^2 / 2$; $dt_E = 6 : 10$. Let the allowed maximum waiting time be 15 min and the maximum total number of transfers $N^+ = 2$ and other parameters the same as [19] and all other user-defined parameters are set equal to 1.0 to calculate the attractive connection set between the OD pair from A-Village to X-City. The convergence tolerance gap is set equal to 0.01.

The assignment was run on a PIII/3.0MHz, 1 GB Ram computer. The computation time is approximately 1 min. Figure 3 illustrates that for $\theta = 1.0$, the gap function converges very rapidly in the beginning and shows small fluctuation after 10 iterations. The MSA algorithm converges very rapidly. Moreover, the final gap function of the converged solution is less than 0.1%, which indicates that the solution is sufficiently close to the equilibrium solution. On the other hand, Figure 4 illustrates that the flows assigned to the five connections have small fluctuations after 70 iterations, suggesting the good convergence property of this MSA algorithm. Each connection has got an excellent flow solution for the overload delay.

The resultant connection flows and their overload delays with $\gamma = 1.0$ are shown in Tables 2 and 3 for $\theta = 0.1, 1.0, 2.0$, and 5.0.

It can be seen in Table 2 that, with capacity constraints, the total flow of c_1 and c_4 could not exceed the vehicle capacity, 200 persons, because these two connections use the same vehicle of bus line 1, which departs at 6:10, visits station at 6:22, and then arrives at X-city at 6:55. The same thing happens among c_2 and c_5 . The flow of c_1 increases as θ increases. For the uncongested connection, the flow for c_5 decreases as θ increases because the passenger perceived a reduced travel cost. For congested connections, in spite of the capacity deficiency, their flows still increase, in other words, in the situation of vehicle capacity deficiency, people

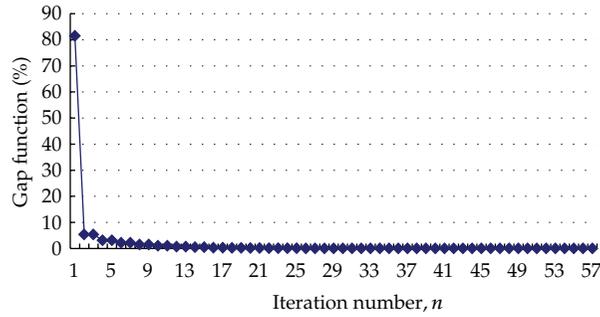


Figure 3: Convergence characteristics of assigned flow for $\theta = 1.0$.

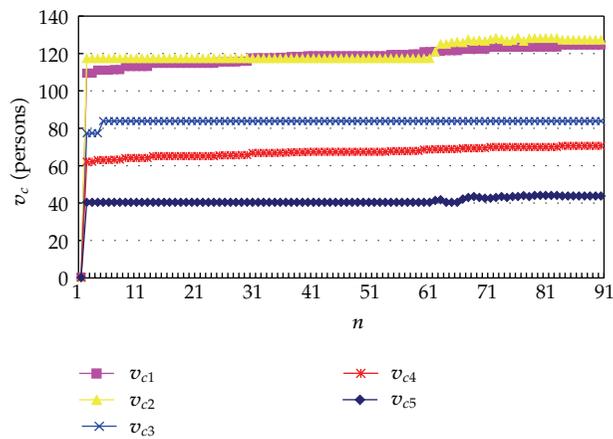


Figure 4: Assigned flows the five connections for $\theta = 1.0$.

Table 2: The resultant connection flows for various θ in comparison with UE result, in persons.

θ	Connections				
	c_1	c_2	c_3	c_4	c_5
0.1	94.57	93.10	89.30	89.35	83.68
1.0	124.47	127.24	83.88	70.62	43.79
2.0	147.82	163.93	71.24	47.59	19.42
5.0	176.17	190.13	50.00	23.83	9.87
UE	200.00	200.00	50.00	0.00	0.00

Table 3: The resultant connection overload delay for various θ , in minute.

θ	Connections				
	c_1	c_2	c_3	c_4	c_5
0.1	0.00	0.00	0.00	0.00	0.00
1.0	24.15	0.00	0.00	9.15	0.00
2.0	32.32	2.21	0.00	17.32	0.00
5.0	33.43	5.32	0.00	26.06	0.00

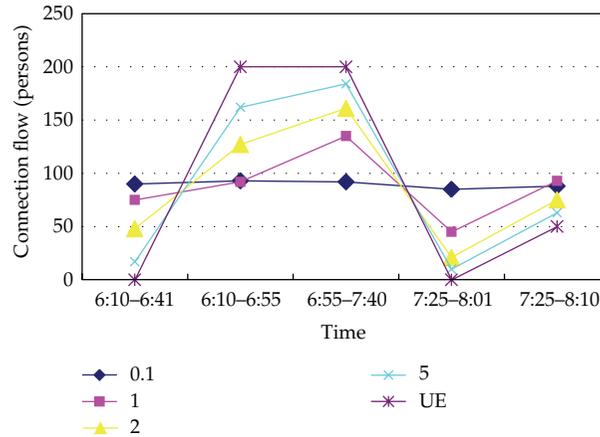


Figure 5: Passenger behaviour in choosing the departure time and the connection for various θ .

would still get on the congested vehicles even though they know the routes and the vehicle schedule clearly. The reason might be, they know clearly that waiting for the successor connection would not be able to reduce the total cost anyway. Consequently, they would rather catch the first incoming vehicle until its capacity runs out. When $\theta = 5.0$, the stochastic connection flows are close to that of UE pattern.

It can be seen in Table 3 that the overload delay of congested connection ($c_1 \sim c_4$) would increase with the increase of θ . This is because passengers have more information about the route and the vehicle schedule and are thus clear that there would be no way to find another connection with smaller cost between origin and destination. As a result everyone would get on the first arriving bus until it becomes full, which would result in an increase in congestion delay.

For various θ , the passenger behaviour embodied in choosing the departure time, and the connection is shown in Figure 5. We can clearly see that for $\theta = 0.1$, passengers have limited information about the network condition, so that they would choose their departure time evenly. On the other hand, when they have more knowledge about the connections and the schedule, they would choose the connection vehicle which would result in a smaller total cost. For instance, less passengers would choose to board the vehicle of this connection (c_5) if they know more (increase of θ) about the schedule at time interval 7:25~8:01.

6. Conclusions

In this paper, a SUE assignment model is proposed for schedule-based transit networks with vehicle capacity constraints. A solution algorithm is developed. The stochastic effects of the passenger's behavior and vehicle timetable, vehicle capacity are incorporated in the model.

The attractive set problem which is conventionally considered only for frequency-based transit networks is formulated for the schedule-based transit network. The generalized cost model is set to determine the costs of connections between OD pairs. We also analyzed a mathematical programming problem equivalent to the SUE assignment problem in schedule-based transit networks with capacity constraints. When a connection segment reaches its capacity level, it is proven that the Lagrange multipliers of the mathematical problem give the equilibrium passenger overload delays in this transit network.

Passenger overload delay is determined endogenously by the equilibrium characteristics and vehicle capacity of the schedule-based transit network in addition to the cost functions of each link used in the existing approaches. The overload delay varies with passenger's knowledge about the scheduled time and the transit lines.

The model proposed in this paper was applied to one special situation for congested transit networks. We found that greater knowledge of the capacity overrun path would result in more overload delay, but with little time to get to the destinations. For future research, several extensions of this model are possible and have the potential to enrich the models available for transit planning such as a SUE assignment with elastic transit demand, SUE assignment for multiple classes of passengers, and dynamic SUE assignment for transit systems.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant no. 41071101 and no. 51208445), and by the Fundamental Research Funds for the Central Universities (Grant no. 2010121074).

References

- [1] A. Babazadeh and H. Z. Aashtiani, "Algorithm for equilibrium transit assignment problem," *Transportation Research Record*, no. 1923, pp. 227–235, 2005.
- [2] R. B. Dial, "Transit pathfinder algorithms," *Highway Research Record*, vol. 205, pp. 67–85, 1967.
- [3] R. B. Dial, "A probabilistic multipath traffic assignment model which obviates path enumeration," *Transportation Research*, vol. 5, no. 2, pp. 83–111, 1971.
- [4] K. Fearnside and D. P. Draper, "Public transport assignment. A new approach," *Traffic Engineering and Control*, vol. 13, no. 7, pp. 298–299, 1971.
- [5] C. Chriqui and P. Robillard, "Common bus lines," *Transportation Science*, vol. 9, no. 2, pp. 115–121, 1975.
- [6] CH. J. De Cea, J. Enrique, and L. Fernandez, "Transit assignment to minimal routes: An efficient new algorithm," *Traffic Engineering and Control*, vol. 30, no. 10, pp. 491–494, 1989.
- [7] H. Spiess, *Contribution à la théorie et aux outils de planification des réseaux de transport urbains [Ph.D. thesis]*, Département d'Informatique et Recherche Opérationnelle, Montreal, Canada, 1984.
- [8] H. Spiess and M. Florian, "Optimal strategies: a new assignment model for transit networks," *Transportation Research Part B*, vol. 23, no. 2, pp. 83–102, 1989.
- [9] J. H. Wu, M. Florian, and P. Marcotte, "Transit equilibrium assignment: a model and solution algorithms," *Transportation Science*, vol. 28, no. 3, pp. 193–203, 1994.
- [10] M. Gendreau, *Etude approfondie d'un modèle d'équilibre pour l'affectation de passagers dans les réseaux de transports en commun [Ph.D. thesis]*, Département d'Informatique et Recherche Opérationnelle, Montreal, Canada, 1984.
- [11] W. H. K. Lam, Z. Y. Gao, K. S. Chan, and H. Yang, "A stochastic user equilibrium assignment model for congested transit networks," *Transportation Research Part B*, vol. 33, no. 5, pp. 351–368, 1999.
- [12] R. Cominetti and J. Correa, "Common-lines and passenger assignment in congested transit networks," *Transportation Science*, vol. 35, no. 3, pp. 250–267, 2001.
- [13] F. Kurauchi, M. G. H. Bell, and J.-D. Schmöcker, "Capacity constrained transit assignment with common lines," *Journal of Mathematical Modelling and Algorithms*, vol. 2, no. 4, pp. 309–327, 2003.
- [14] S. Nguyen, S. Pallottino, and M. Gendreau, "Implicit enumeration of hyperpaths in a logit model for transit networks," *Transportation Science*, vol. 32, no. 1, pp. 54–64, 1998.
- [15] J. D. Schmöcker, *Dynamic capacity constrained transit assignment [Ph.D. thesis]*, Centre for Transport Studies, Department for Civil and Environmental Engineering, Imperial College London, London, UK, 2006.
- [16] M. Florian, "Deterministic time table transit assignment," in *Paper of EMME/2 Users Group Meeting*, Shanghai, China, 1999.

- [17] C. O. Tong and S. C. Wong, "A stochastic transit assignment model using a dynamic schedule-based network," *Transportation Research, Part B*, vol. 33, no. 2, pp. 107–121, 1998.
- [18] C. O. Tong and S. C. Wong, "Schedule-based time-dependent trip assignment model for transit networks," *Journal of Advanced Transportation*, vol. 33, no. 3, pp. 371–388, 1999.
- [19] M. Friedrich and S. Wekeck, "A schedule-based transit assignment model addressing the passengers' choice among competing connections," in *Proceedings of the Conference on Schedule-Based Approach in Dynamic Transit Modeling: Theory and Applications*, pp. 159–173, Ischia, Italy, 2002.
- [20] M. Friedrich, "Multi-day dynamic transit assignment," in *Proceedings of the Conference on Schedule-Based Dynamic Transit modeling (SBDTM '05)*, Ischia, Italy, 2005.
- [21] O. A. Nielsen, "A stochastic transit assignment model considering differences in passengers utility functions," *Transportation Research Part B*, vol. 34, no. 5, pp. 377–402, 2000.
- [22] O. A. Nielsen and R. D. Frederiksen, "Optimisation of timetable-based, stochastic transit assignment models based on MSA," *Annals of Operations Research*, vol. 144, pp. 263–285, 2006.
- [23] A. Nuzzolo, "Schedule-based transit assignment models," in *Advanced Modeling for Transit Operations and Service*, H. K. William Lam and G. H. Michael Bell, Eds., Pergamon, 2003.
- [24] A. Nuzzolo and U. Crisalli, "The Schedule-Based approach in dynamic transit modelling: a general overview," in *Schedule-Based Dynamic Transit Modeling, Theory and Applications*, N. H. M. Wilson and A. Nuzzolo, Eds., Kluwer Academic Publishers, 2004.
- [25] W. Xu, S. He, R. Song, and S. S. Chaudhry, "Finding the K shortest paths in a schedule-based transit network," *Computers & Operations Research*, vol. 39, no. 8, pp. 1812–1826, 2012.
- [26] J. de Cea and E. Fernandez, "Transit assignment for congested public transport systems: an equilibrium model," *Transportation Science*, vol. 27, no. 2, pp. 133–147, 1993.
- [27] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.
- [28] M. Chen and A. S. Alfa, "Algorithms for solving fisk's stochastic traffic assignment model," *Transportation Research Part B*, vol. 25, no. 6, pp. 405–412, 1991.
- [29] M. G. H. Bell, W. H. K. Lam, G. Ploss, and D. Inandi, "Stochastic user equilibrium assignment and iterative balancing," in *Proceedings of the 12th International Symposium on Transportation and Traffic Theory*, Berkeley, Calif, USA, 1993.
- [30] M. G. H. Bell, "Stochastic user equilibrium assignment in networks with queues," *Transportation Research Part B*, vol. 29, no. 2, pp. 125–137, 1995.

Research Article

Crossing at a Red Light: Behavior of Cyclists at Urban Intersections

**Xiaobao Yang,¹ Mei Huan,¹ Bingfeng Si,¹
Liang Gao,¹ and Hongwei Guo²**

¹ MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology,
Beijing Jiaotong University, Beijing 100044, China

² Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Xiaobao Yang, yangxb@bjtu.edu.cn

Received 10 June 2012; Revised 23 August 2012; Accepted 30 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Xiaobao 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.

To investigate the relationship between cyclist violation and waiting duration, the red-light running behavior of nonmotorized vehicles is examined at signalized intersections. Violation waiting duration is collected by video cameras and it is assigned as censored and uncensored data to distinguish between normal crossing and red-light running. A proportional hazard-based duration model is introduced, and variables revealing personal characteristics and traffic conditions are used to describe the effects of internal and external factors. Empirical results show that the red-light running behavior of cyclist is time dependent. Cyclist's violating behavior represents positive duration dependence, that the longer the waiting time elapsed, the more likely cyclists would end the wait soon. About 32% of cyclists are at high risk of violation and low waiting time to cross the intersections. About 15% of all the cyclists are generally nonrisk takers who can obey the traffic rules after waiting for 95 seconds. The human factors and external environment play an important role in cyclists' violation behavior. Minimizing the effects of unfavorable condition in traffic planning and designing may be an effective measure to enhance traffic safety.

1. Introduction

Urban traffic problem is widely recognized as one of the main maladies of life in large cities. Many scholars have paid much attention to the traffic problem [1, 2]. A Mix of non-motorized and motorized vehicles is an important traffic type in China. Some surveys show that the non-motorized vehicle is one of the most widely used traffic tools in Chinese daily travel activity [3]. At present, cycling still has heavy proportion among all travel modes in China, and in Tianjin, as large as over 60% [4]. Even in the developed countries, bicycle travel is recognized as a low-energy consumption and as being healthy to the users and it does not damage the health of others. Meanwhile, the electric bike (e-bike) has emerged as a popular mode of

transportation in many large cities during recent years. E-bike use has rapidly expanded in China, in the process changing the mode split of many cities. Currently, China produces over 20 million e-bikes yearly, up from a few thousands a decade ago [5]. E-bikes in China are defined as electric two-wheelers with relatively low speeds and weights compared to a motorcycle. Both bicycle-style e-bikes (with functioning pedals) and scooter-style e-bikes (with many features of gasoline scooters) are classified as bicycles and are given access to bicycle infrastructure.

However, the growing popularity of cycling traffic also entails safety concerns as observed in accident statistics. Accident analysis reveals that over 60% of fatal crashes involving cyclists result from violation of traffic rules [6]. One typical type of rule violation behavior is violation behavior at red period. Because of the poor law enforcement and people's low safety awareness, violation behavior at red period is rather prevalent and represents a substantial safety problem in Chinese urban intersections. Especially, electric bicycles with relatively high speed are likely to increase the risk of traffic incident.

Previous research on drivers and pedestrians also points to several variables of interest regarding violation behaviors. Keegan and O'Mahony gave reports about pedestrians' street-crossing behavior influenced by travel distance and waiting time [7]. Other researchers paid much attention to the influences of personal features on the street-crossing behavior [8–10]. Some useful reviews of the existing research on pedestrian street-crossing behavior in urban areas can be found in Ishaque and Noland [11] and Papadimitriou et al. [12].

Unfortunately, only a few studies have investigated the violation behavior of cyclists. Johnson et al. identified three distinct types of violated cyclists that are exposed to different levels of risk: racers, impatient, and runners [13]. Johnson et al. used video recordings to analyze urban commuter cyclists' violation behaviors in Melbourne, VIC, Australia [14]. The field observation approach was also used by researchers studying the influence of cycle paths on accident numbers [15].

In this paper, a hazard-based duration approach is adopted to describe the cyclist violation behavior at signalized intersections. The hazard-based duration models have been used extensively in biometrics and reliability engineering for decades [16]. Duration models can be used to determine causality in duration data and they are also useful tools in the field of transportation [17–21]. These models represent a type of analytical methods to describe the duration of a certain state and how various factors have affected the duration. More importantly, duration models can deal with not only uncensored data but censored data. For example, the exact waiting duration reflecting cyclist endurance cannot be observed if cyclists could wait until the permission of traffic rules. However, most statistical models are unable to analyze these uncensored data. Accordingly, cyclists' waiting times are modeled by a proportional hazard-based duration model. The covariates relevant to traffic conditions and personal features are investigated to capture the influenced factors of cyclist behavior. The results give the time when cyclists are easy to violate traffic rules and the significantly influential factors on waiting behavior. The findings will imply some effective countermeasures for improving the road safety of urban intersections.

2. Method

2.1. Duration Model

The variable of interest in duration model is the survival time that elapsed from the beginning of an event until its end. The waiting time of a cyclist at red light can be regarded as

the waiting duration that starts when a cyclist arrives at the intersection at the red period and ends when the cyclist begins to cross the intersection.

Let T be a nonnegative variable representing the waiting duration of a cyclist at a signalized intersection. Let $f(t)$ denote the probability density function of T and the cumulative distribution be

$$F(t) = \Pr(T \leq t) = \int_0^t f(u) du. \quad (2.1)$$

Let $S(t)$ denote the probability that the waiting duration does not end prior to t

$$S(t) = 1 - F(t) = \Pr(T > t) = \int_t^{\infty} f(u) du. \quad (2.2)$$

$S(t)$ is called survival function or survivor probability. The survival function is defined to be the probability that the waiting time of a cyclist at a red light is longer than some specific time, t .

In the survival analysis, T can be characterized by a hazard function, $h(t)$. The hazard function under study is the instantaneous rate at which the waiting duration will end in an infinitesimally small time period, Δt , after time t , given that the duration time has lasted to time t

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t \times \Pr(T \geq t)} = \frac{f(t)}{S(t)} = \frac{-d \ln S(t)}{dt}. \quad (2.3)$$

Note that the waiting time of a cyclist is influenced by various factors. The influential factors can be defined as a vector of explanatory variables, $\mathbf{x} = (x_1, x_2, \dots, x_p)'$. To accommodate the effects of these influential factors is a main objective of this paper. Thus the proportional hazard form is introduced, which specifies the effects of explanatory variables to be multiplicative on a hazard function

$$h(t) = h_0(t)g(\mathbf{x}, \boldsymbol{\beta}), \quad (2.4)$$

where $h_0(t)$ is called the baseline hazard function and can be interpreted as the hazard function when all covariates are ignored. $g(\cdot)$ is a known function to represent the effects of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ is a vector of estimable coefficients for \mathbf{x} . In this paper, a typical specification with $g(\mathbf{x}, \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}\mathbf{x})$, which was proposed by Cox [22], is used. The specification is convenient since it guarantees the positivity of the hazard function without placing constraints on the signs of the elements of $\boldsymbol{\beta}$. The Cox proportional hazard model is

$$h(t) = h_0(t) \exp(\boldsymbol{\beta}\mathbf{x}). \quad (2.5)$$

Combining (2.3) and (2.5), the survival function can be written as

$$S(t) = \exp\left[-\int_0^t h(w)dw\right] = \left\{\exp\left[-\int_0^t h_0(w)dw\right]\right\}^{\exp(\beta\mathbf{x})} = \{\exp[-H_0(t)]\}^{\exp(\beta\mathbf{x})}, \quad (2.6)$$

where $H_0(t) = \int_0^t h_0(w)dw$ represents the baseline cumulative hazard function. Thus, the covariates can be incorporated into the survival function.

2.2. Model Estimation

The main interest of this paper is to identify from the p covariates a subset of variables that affects the hazard more significantly, and consequently, the waiting duration time at a signalized intersection. We are concerned with the regression coefficients. If β_i is zero, the corresponding covariate is not related to the waiting time. If β_i is not zero, it represents the magnitude of the effect of x_i on hazard when the other covariates are considered simultaneously.

To estimate the coefficients, $\beta_1, \beta_2, \dots, \beta_p$, a partial likelihood method is adopted. Suppose that k of the duration times from n cyclists is observed and distinct. Let $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ be the ordered k distinct duration times with corresponding covariates $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(k)}$. Let $\mathbf{R}(t_{(i)})$ be the risk set at time $t_{(i)}$. $\mathbf{R}(t_{(i)})$ consists of all cyclists whose duration times are at least $t_{(i)}$. For the particular duration time $t_{(i)}$, conditionally on the risk set $\mathbf{R}(t_{(i)})$, the probability is

$$\frac{\exp\left(\sum_{j=1}^p b_j x_{j(i)}\right)}{\sum_{l \in \mathbf{R}(t_{(i)})} \exp\left(\sum_{j=1}^p b_j x_{jl}\right)} = \frac{\exp(\beta\mathbf{x}_{(i)})}{\sum_{l \in \mathbf{R}(t_{(i)})} \exp(\beta\mathbf{x}_l)}. \quad (2.7)$$

Each distinct duration time contributes a factor and hence the partial likelihood function is

$$L(\beta) = \prod_{i=1}^k \frac{\exp\left(\sum_{j=1}^p b_j x_{j(i)}\right)}{\sum_{l \in \mathbf{R}(t_{(i)})} \exp\left(\sum_{j=1}^p b_j x_{jl}\right)} = \prod_{i=1}^k \frac{\exp(\beta\mathbf{x}_{(i)})}{\sum_{l \in \mathbf{R}(t_{(i)})} \exp(\beta\mathbf{x}_l)} \quad (2.8)$$

and the log-partial likelihood is

$$l(\beta) = \log L(\beta) = \sum_{i=1}^k \left\{ \beta\mathbf{x}_{(i)} - \log \left[\sum_{l \in \mathbf{R}(t_{(i)})} \exp(\beta\mathbf{x}_l) \right] \right\}. \quad (2.9)$$

The overall goodness of fit of the model estimation is determined by the likelihood ratio (LR) statistics, which is specified as

$$X_L = -2 \left[l(\beta_0) - l(\hat{\beta}) \right], \quad (2.10)$$

Table 1: Covariates selection and explanation.

Covariate	Type	Explanation
AG (age group)	Binary indicator	1 if old (≥ 50), 0 otherwise
GEN (gender)	Binary indicator	1 if male, 0 female
NT (nonmotor vehicle type)	Binary indicator	1 if electric bike, 0 human-powered bike
WN (waiting number)	Continuous variable	The number of other cyclists that are waiting for a green light when arrives
CN (crossing number)	Continuous variable	The number of other cyclists that are crossing against the red light when arrives
TC (twice crossing)	Binary indicator	1 if twice crossing behavior, 0 otherwise
MV (motor vehicle volume)	Continuous variable	Average motor vehicle volume per lane per min on red-light phase when the cyclist arrives
TT (travel time)	Binary indicator	1 if a cyclist travels in peak hour, 0 otherwise

where $l(\beta_0)$ is the log-partial likelihood for null model with all the regression coefficients are set as zero and $l(\hat{\beta})$ is the log-partial likelihood at convergence with p regression coefficients. The Cox proportional hazard model has been widely cited in the literature. For the estimation of $H_0(t)$ and other detailed discussion of this model see, Lee and Wang [16] and Bhat [17].

2.3. Covariate Selection

The covariate selection takes into account the previous research [13, 14] and arguments regarding the effects of the exogenous variables on cyclist crossing behavior. The practical effects on waiting behavior and the feasibility of data acquisition are considered in the covariate selection. Two broad sets of variables are considered as covariates: personal characteristics and traffic conditions. Personal characteristics involve age and gender. The selected covariates of traffic conditions can determine the effects on the waiting time and traffic volume. The following covariates, as shown in Table 1, are adopted to construct the duration model.

3. Survey and Data

3.1. Site Survey Design

To record cyclists' waiting durations, the whole red-light period of a signal cycle was observed as a data collection unit. Only the cyclists who arrived in the red-light period were defined as a valid sample. The waiting duration was from the time a cyclist arrived at the crossing location to the time he/she began to cross. It can be classified into two kinds: uncensored data and censored data. The uncensored data was defined as the waiting duration which ended within the red-light period (violating crossing). Otherwise, the waiting duration was called as the censored data as long as it ended within the green-light period (normal crossing). For the censored data, it is unknown about the exact waiting duration which can reflect the endurance of waiting time for cyclists.

The site survey was conducted at three selected signalized intersections near Jiaotong University in Beijing, China. Data collections were done by placing video cameras at each location. The survey periods included peak hour (7:30 a.m.–9:30 a.m.) and offpeak hour

Table 2: Estimation in waiting duration model.

Variable	Coefficient (β)	Standard error	Wald statistic	Exp (β)	<i>P</i> value
AG	-0.673	0.395	2.905	0.510	0.088
GEN	0.325	0.227	2.053	1.383	0.152
NT	0.471	0.193	5.953	1.601	0.015
WN	-0.129	0.031	17.106	0.879	<0.001
CN	0.342	0.071	23.418	1.408	<0.001
TC	0.853	0.217	15.453	2.346	<0.001
MV	-0.170	0.065	6.876	0.844	0.009
TT	0.365	0.124	8.595	1.440	0.003

(10:00 a.m.–4:00 p.m.). The survey area covered the zebra crossing and a part of traffic lanes so that the cyclist crossing behavior and the corresponding traffic conditions can be monitored clearly. Some additional explanations are needed for the site survey.

- (a) The signals were old traditional person heads so that the influence of type of signals could be neglected [7]. The selected sites had similar characteristics involved geometric, and traffic conditions, traffic control.
- (b) The survey was conducted in good weather and the absence of pointsmen. Cyclists were unobtrusively observed.

3.2. Descriptive Statistics

Of the 459 valid observations, 295 (64.27%) cyclists violated the traffic regulations. The average waiting time of all samples was 25.16 seconds, with a standard deviation of 27.13 seconds. The average waiting time of the violating crossing was 15.71 seconds while the average waiting time of the normal crossing is 43.14 seconds. The maximum waiting duration was 116 seconds while the minimum was 0 second. The latter means people cross the street without any wait. This descriptive statistic cannot reflect the exact waiting behavior due to the neglect of the censored data. The estimation of the waiting duration with censored data will be discussed later.

4. Empirical Results

The results are discussed in two sections. The overall results are presented in the first section including model fit statistics and survival probability estimation. The second subsection presents the effects of covariates.

4.1. Overall Results

(1) Model fit statistics: the LR statistic of the estimated model clearly indicates the overall goodness of fit (the LR statistic is 3201.0, which is greater than the chi-squared statistic with 8 degrees of freedom at any reasonable level of significance). The significant level corresponding to each covariate is given by *P* value in Table 2. From the results, most of the included covariates are statistically significant at the 0.10 level of significance. It means that these covariates are significantly related to violation behavior. Only gender has relative low significant level. It is partly because that the female rate (24.2%) in the sample is relative

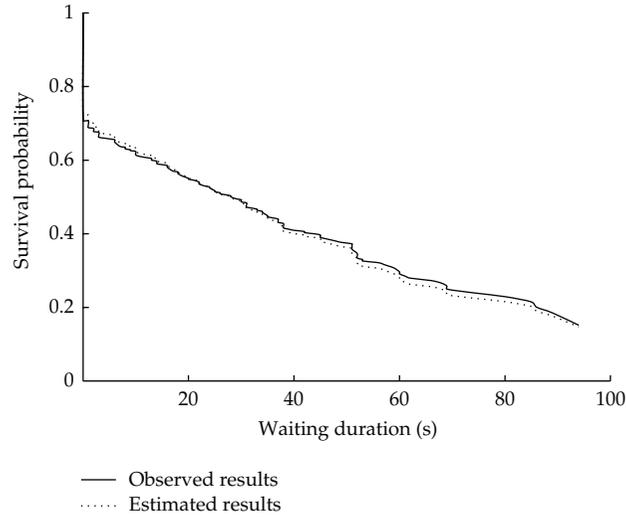


Figure 1: Survival probabilities versus waiting duration.

low. The significance level of each covariate suggests that the importance of covariate should be interpreted carefully.

(2) Survival probability: Figure 1 gives the survival probability calculated by the duration model, which represents the probability of complying with the traffic rules while waiting at the signalized intersections. The survival probability for estimated model presents a general decline trend with elapsed waiting duration. The survival probability can be divided into three parts according to the gradient. Firstly, a sharp decline for the short duration indicates that there are a number of cyclists would violate to cross without any delay. Especially, about 32 percent of cyclists can be defined as risk takers since they show high violation inclination and very low waiting endurance (<3 seconds). Then, the probability decreases smoothly from 3 seconds to 95 seconds. This steady reduction reflects that the number of cyclist violations is increasing continuously. The declining trend of the survival probability indicates that the red-light running behavior of most cyclists is time dependent. It means that cyclists are easy to end waiting duration and violate the traffic rules with the elapsed duration. Note that about half of the observed cyclists cannot endure 29 seconds or longer. Finally, there are 15 percent of cyclists who wait and wait longer, and they are generally non-risk takers.

(3) Figure 1 also gives the comparison between the estimated survival probability and the observed survival probability. Here, the estimated results are calculated by the Cox proportional hazard model; while the observed results are calculated by the nonparametric approach in which the covariate effects are not considered. The detailed discussion of the non-parametric approach can refer to the work of Lee and Wang [16]. The results show that there are some differences between them though the general shape is the same between the two results. Specifically, compared to the estimated results, the observed survival probability is smaller until about 24.0s, larger thereafter. This difference is expected to be the covariate effects, at least partly. The observed results indicate the waiting time under the specific condition for individual sample, while the estimated results indicate the waiting time under the average condition for all the samples. The estimated survival probability reflects the

characteristics of the waiting time which has an average value for every variable. Any change of the variables could influence the estimated results. The effects of variables are discussed in the next subsection.

4.2. Analysis of Covariate Effects

According to (2.5), the effects of the explanatory variables can be interpreted by the signs of the coefficients in a rather straightforward fashion. If the coefficient is negative, it implies that an increase in the corresponding variable decreases the hazard rate, or equivalently, increases the waiting duration. With regard to the magnitude of the variable effects, when a variable changes by one unit, the hazard would change by $[\exp(\beta) - 1] \times 100\%$.

To assess the effects of the explanatory variables on the duration time, a function of hazard ratio (HR) can be obtained by dividing both sides of (2.5) by $h_0(t)$, yield

$$\log \frac{h_i(t)}{h_0(t)} = \boldsymbol{\beta} \mathbf{x}_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}, \quad (4.1)$$

where the x 's are covariates for the i th cyclist; $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ is a vector of the coefficients which has been estimated by using the Cox proportional hazard model. The left side of (4.1) is a function of hazard ratio (HR) and the right side is a linear function of the covariates and their respective coefficients. The HR can represent the multiple relations between the hazard under the covariate effects and the hazard when all variables are ignored.

If the covariates are standardized about the mean and the model used is

$$\log \frac{h_i(t)}{h_0(t)} = \boldsymbol{\beta}(\mathbf{x}_i - \bar{\mathbf{x}}) = \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + \cdots + \beta_p(x_{pi} - \bar{x}_p), \quad (4.2)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ and \bar{x}_j is the average of the j th covariate for all cyclists, the left side of (4.2) is the logarithm of the relative hazard ratio (RHR). RHR represents the hazard ratio for a cyclist with a given set of values to that for a cyclist which has an average value for every covariate. If RHR is more than one, it means that the covariate effects can increase the hazard and so the variables are favorable. That is to say, the waiting time in such a favorable condition is less than the average level of the survey sample. On the contrary, the unfavorable variable corresponds to a low hazard. Therefore, a cyclist in the unfavorable condition would have longer waiting time than that in the favorable condition.

In order to make a quantitative analysis on the effects of covariates, the relative hazard for each variable is calculated by considering favorable or unfavorable values of that variable, assuming that other variables are at their average value. The favorable or unfavorable values of that variable are given on the basis of the hazard with the value of the variable. The value of the variable with the low hazard is regarded as the favorable condition. Take the age as an example, old people are defined as the favorable condition since old people have lower violation risk than young people. The assumed conditions and corresponding RHRs and HRs are shown in Table 3. The RHRs for three continuous covariates are shown in Figure 2.

The effect of age (AG) indicates that older cyclists have longer waiting time. This is partly because older cyclists have stronger risk consciousness of traffic violations. In addition, older cyclists' trip purposes are seldom related to work or school so they are not in a hurry.

Table 3: Estimation of RHRs and HRs for assumed covariates.

Variable	Mean	Variables value		Relative hazard ratio		Hazard ratio
		Favorable	Unfavorable	Low hazard	High hazard	
AG	0.758	1	0	0.85	1.67	1.96
GEN	0.075	0	1	0.98	1.35	1.38
NT	0.535	0	1	0.78	1.24	1.60
WN	1.913	8	0	0.35	1.28	2.81
CN	0.235	0	4	0.92	3.62	3.93
TC	0.059	0	1	0.95	2.23	2.35
MV	6.610	8	1	0.79	2.60	3.29
TT	0.501	0	1	0.83	1.20	1.44

Note that this conclusion is a statistical result; traffic violations involved older cyclists are also common sometimes.

The effect of gender (GEN) indicates that male cyclists have shorter waiting time and higher tendency to disobey the traffic rules. They are 1.38 times more likely than females to have shorter waiting times. Hamed reported that male pedestrians are 2.61 times more likely than females to have shorter waiting times [23], and other qualitatively similar results were obtained by Tiwari et al. [24].

The effect of covariate nonmotor vehicle type (NT) indicates that cyclists of electric bike have shorter waiting time and higher tendency to disobey the traffic rules. They are 1.60 times more likely than human-powered bicyclists to have shorter waiting times.

The waiting time of cyclists would increase with the bigger number of other cyclists that are waiting for a green, (WN) when arrives (see Figure 2(a)). Otherwise, the waiting time decreases with the bigger number of other cyclists that are crossing against the red light (CN) when arrives (see Figure 2(b)). This is caused by two reasons. First, many people may consider that the more people cross together, the safer they would be. They take it for granted that drivers must yield to a group of people more often than one person. Second, the conformity psychology would work well in dense cyclist environments.

The effect of covariate TC (twice crossing) shows that the cyclists of twice crossing have higher hazard and shorter waiting time. Cyclists who are apt to twice crossing behavior have little or no patience to wait at a red light. They are 2.35 times more likely than one-time crossing cyclists to have shorter waiting times.

The effect of covariate MV (motor vehicle volume) indicates that heavy traffic can increase waiting time or decrease the risk of cyclist violations (see Figure 2(c)). This is because that the larger motor vehicle volume is, the smaller the average time gap between successive cars is.

The characteristics of travel time also have impacts on cyclists' red-light running behavior. It indicates that cyclists are at high risk level of traffic violation in peak hour. The cyclists who travel in peak hour are 1.44 times more than those who travel in offpeak hour to end waiting duration and cross illegally. In peak hour period, both cyclists and drivers are in a hurry to the destination related to work or school, so the heavy mixed traffic with impatient cyclists and drivers would cause traffic accidents easily.

5. Conclusions

This paper applies a proportional hazard-based duration model to study the cyclist crossing behavior at signalized intersections by using data acquired in Beijing, China. The crossing

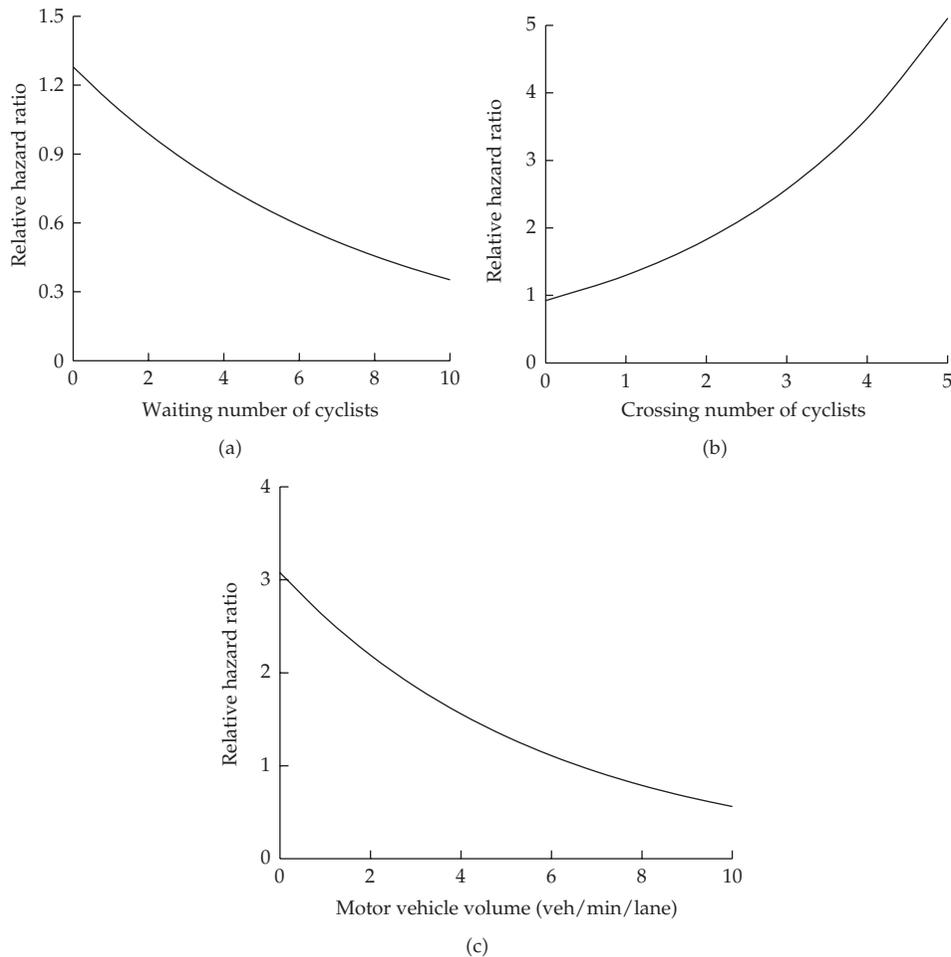


Figure 2: Relative hazard ratios for three continuous variables.

behavior is examined by modeling the duration between the arrival at survey area and the start to cross the intersection. If cyclists violate the traffic rules to cross the intersection, their waiting times are recorded as uncensored data, while the waiting durations of normal crossing are recorded as censored data.

The paper provides several important insights into the determinants of the regularity and frequency of cyclist crossing behavior, especially the relation between violation behavior and waiting duration. First, the results indicate that the crossing behavior of cyclists is time dependent, as well as the risk of traffic violation. Cyclists' crossing behavior presents positive duration dependence, which also implies a "snowballing" effect. It means the longer the time has elapsed since the start of the waiting duration, the more likely cyclists will end the wait soon. Such positive duration dependence also indicates that longer waiting time would increase the risk of cyclist violation. Second, some crucial time points deserve our concern: 3 seconds and 95 seconds. The 3 seconds indicate cyclists who are at high risk of violating crossing the street and low waiting time, and they account for 32% of the sample in the study. The duration of 95 seconds reflects the cyclists' endurance. About 15 percent of all

the cyclists can obey the traffic rules after waiting for 95 seconds. These people are generally non-risk takers. Third, the human factors and the external environment play an important role in red-light running behavior. Various factors in the unfavorable condition could increase the risk of traffic violation, as well as traffic accidents. The effects of covariates can help to modify cyclists' crossing behavior. Specifically, rational traffic planning and designing should fully consider cyclist behavioral characteristics. More importantly, minimizing the effects of unfavorable condition involved human factors may be an effective measure to obtain conscious cooperation and behavioral changes of cyclists. Finally, it is noted that, for different cities, the model should be estimated by using the specified field data. Additionally, the explanatory variables can be chosen flexibly according to the research aim and the traffic reality.

In terms of the future work, more parameters under different situations should be taken into account. Next, some engineering solutions should be proposed to improve the safe crossing behavior of cyclists in urban traffic environment. In addition, from the viewpoint of cyclist prevention, the interaction between cyclists and motor vehicles could be analyzed based on such crossing behavior. Findings from this paper may partly supplement previous research which helps us in inspiration. It is also hoped that these findings may give better understanding of cyclist behavioral characteristics at signalized intersections and help to plan and design proper facilities for non-motorized vehicles.

Acknowledgments

This work was supported by the National Basic Research Program of China (Grant no. 2012CB725400), the National Natural Science Foundation of China (Grant no. 70901005, 71131001), and the Fundamental Research Funds for the Central Universities (Grant no. 2012YJS067).

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] C. K. Wong and Y. Y. Lee, "Convergence study of minimizing the nonconvex total delay using the lane-based optimization method for signal-controlled junctions," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 858731, 14 pages, 2012.
- [3] K. J. Krizek, "Estimating the economic benefits of bicycling and bicycle facilities: an interpretive review and proposed methods," in *Proceedings of the TRB 2004 Annual Meeting*, 2004, CD-ROM 04-2070.
- [4] D. Wang, T. Feng, and C. Liang, "Research on bicycle conversion factors," *Transportation Research A*, vol. 42, no. 8, pp. 1129–1139, 2008.
- [5] J. X. Weinert, C. Ma, X. Yang, and C. R. Cherry, "Electric two-wheelers in China: effect on travel behavior, mode shift, and user safety perceptions in a medium-sized city," *Transportation Research Record*, no. 2038, pp. 62–68, 2007.
- [6] CRTASR, *China Road Traffic Accidents Statistics Report*, Traffic Administration Bureau of China State Security Ministry, Beijing, China, 2004.
- [7] O. Keegan and M. O'Mahony, "Modifying pedestrian behaviour," *Transportation Research A*, vol. 37, no. 10, pp. 889–901, 2003.
- [8] 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.
- [9] K. Pfeffer, H. P. Fagbemi, and S. Stennet, "Adult Pedestrian behavior when accompanying children on the route to school," *Traffic Injury Prevention*, vol. 11, no. 2, pp. 188–193, 2010.

- [10] R. Zhou and W. J. Horrey, "Predicting adolescent pedestrians' behavioral intentions to follow the masses in risky crossing situations," *Transportation Research F*, vol. 13, no. 3, pp. 153–163, 2010.
- [11] M. M. Ishaque and R. B. Noland, "Behavioural issues in pedestrian speed choice and street crossing behaviour: a review," *Transport Reviews*, vol. 28, no. 1, pp. 61–85, 2008.
- [12] E. Papadimitriou, G. Yannis, and J. Golias, "A critical assessment of pedestrian behaviour models," *Transportation Research F*, vol. 12, no. 3, pp. 242–255, 2009.
- [13] M. Johnson, J. Charlton, and J. Oxley, "Cyclists and red lights: a study of behavior of commuter cyclists in Melbourne," in *Proceedings of the Australasian Road Safety Research, Policing and Education Conference*, 2008.
- [14] M. Johnson, S. Newstead, J. Charlton, and J. Oxley, "Riding through red lights: the rate, characteristics and risk factors of non-compliant urban commuter cyclists," *Accident Analysis and Prevention*, vol. 43, no. 1, pp. 323–328, 2011.
- [15] R. O. Phillips, T. Bjørnskau, R. Hagman, and F. Sagberg, "Reduction in car-bicycle conflict at a road-cycle path intersection: evidence of road user adaptation?" *Transportation Research F*, vol. 14, no. 2, pp. 87–95, 2011.
- [16] E. T. Lee and J. W. Wang, *Statistical Methods for Survival Data Analysis*, John Wiley & Sons, New York, NY, USA, 2003.
- [17] C. R. Bhat, "Duration modeling," in *Handbook of Transport Modelling*, D. A. Hensher and K. J. Button, Eds., pp. 91–111, Elsevier Science, Oxford, UK, 2000.
- [18] C. R. Bhat, T. Frusti, H. Zhao, S. Schönfelder, and K. W. Axhausen, "Intershopping duration: an analysis using multiweek data," *Transportation Research B*, vol. 38, no. 1, pp. 39–60, 2004.
- [19] X. B. Yang, M. Huan, H. W. Guo, and L. Gao, "Car travel time estimation near a bus stop with non-motorized vehicles," *International Journal of Computational Intelligence Systems*, vol. 6, no. 4, pp. 1350–1357, 2011.
- [20] W. Wang, H. Guo, Z. Gao, and H. Bubb, "Individual differences of pedestrian behaviour in midblock crosswalk and intersection," *International Journal of Crashworthiness*, vol. 16, no. 1, pp. 1–9, 2011.
- [21] H. W. Guo, W. H. Wang, W. W. Guo, X. B. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.
- [22] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society B*, vol. 34, no. 2, pp. 187–220, 1972.
- [23] M. M. Hamed, "Analysis of pedestrians' behavior at pedestrian crossings," *Safety Science*, vol. 38, no. 1, pp. 63–82, 2001.
- [24] 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.

Research Article

Successive-Stage Speed Limit on Exit Ramp Upstream of Direct-Type Freeway in China

Hongwei Li,¹ Jian Lu,² Yongfeng Ma,¹ and Yuanlin Liu¹

¹ School of Transportation, Southeast University, Nanjing 210096, China

² School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence should be addressed to Hongwei Li, lihongwei-2008@163.com

Received 18 May 2012; Revised 2 August 2012; Accepted 3 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Hongwei 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 first objective of this study is to analyze a successive-stage speed limit model developed for vehicles along the exit upstream ramp of direct-type freeway in China. This paper (1) explains the necessity to implement speed limit to the exit ramp upstream, (2) analyzes whether speed limit is related to the length of the deceleration lane, vehicle type, saturation, and turning ratio and (3) proposes a speed prediction model and calibrates speed-limit sign validity model and establishes successive-stage speed limit model. *The results.* $\Delta v_{85} \geq 10$ illustrates the necessity of the using speed limit on the exit ramp. Speed-deceleration lane length curve presents two trends bounded by 200 m, so the speed limit should be in accordance with the deceleration length. Speed-small vehicle curve closing to speed-large vehicle curve presents that the vehicle type is not the factor of the speed limit. After curve fitting and polynomial regression, saturation is considered to be the most influential factor of speed. Speed-saturation prediction model and calibrated speed-limit sign validity model are built through linearization. According to the above results, successive-stage speed limit model is established. An exit ramp was implemented to verify the feasibility and validity of the model.

1. Introduction

Ramps provide the connections between freeway and roads and influence traffic efficiency and safety of the freeway and ground roads. In USA, 20% to 30% of freeway truck accidents occur at or near ramps (excluding an additional 10% to 15% that occur at weaving section and surface streets), despite the fact that weaving section account for less than 5% of all freeway lane-miles [1]. Kloeden et al. have provided direct evidence that speeds just 5 km/h above the average in urban (60 km/h) areas, and 10 km/h above average in rural areas, are sufficient to double the risk of a casualty crash [2]. It can be seen that the speed of vehicles, even with minor changes, will have a significant influence on freeway safety. *World Prevent Road Traffic*

Injuries Report [3] pointed out that speed is the first risky influence of collisions. Therefore, controlling vehicle speed and reducing speed dispersion is the key to reduce accident rates on the freeway.

2. Literature Review

According to speed limit determinant factors, speed limit control methods are divided into four categories:

- (1) Road grade and geographic feature: The United States [4] use a legal speed limit method for certain types of road infrastructure. The value of the speed limit is mainly determined according to highway classification and alignment elements. Meanwhile, design speed, operating speed, historical accident records and law enforcement experience, and other factors are considered.
- (2) Δv_{85} : Research on the speed limit on the exit ramp of the freeway has focused on one single limit for a long time. *U.S. Institute of Transportation Engineers* [5] recommended the speed limit to be 5 mph higher than v_{85} (the 85% speed) and accident rates is recommended for making speed limits. Milliken [6] proposed a speed limit model using Δv_{85} (the 85% speed difference) on free flow modified by accident rates. A study in the US [7] mentioned that the speed limit should be obtained by Δv_{85} on the freeway main line and the small vehicle and large vehicle should use different speed limit on the freeway main line. The Poisson specification was applied to characterize the relationship between traffic speeds and crash rates under free-flow conditions in two different areas. The results suggested that the proportion of heavy vehicles is inversely associated with the crash rate, and mean speed contributes to crash rates [8]. Park [9] established the speed limit value model by road alignment, traffic flow, and surrounding environment collected under low accident rate and free traffic flow.
- (3) Driver physiological characteristics: *National cooperative highway research program* [10] studied the relationship between the speed limit, accident rate, and vehicle speed. Data was collected from a questionnaire and traffic accident record. *Georgia Department of Driver Services* [11] evaluated driver's feeling of vehicle speed limit and safety using drivers' tolerance degrees to exceed speeds.
- (4) The comprehensive influential factors: In Australia, ARRB [12] developed a road safety software—XLIMITES with a complex decision-supporting method. The method determines speed limits by speed limit already in use, land type, road characteristics, and historical accidents.

The disadvantages in above mentioned methods are:

- (1) Application scope of legal speed limit method is limited: In some sections, the actual situation on the freeway does not match the range of the legal limit speed and, therefore, legal speed limit method cannot be used under some situation.
- (2) Single speed limit is not related with the speed of change and accident data is not easy to obtain. Solomon [13] found that the number of accidents has a U-curve shape related with $V_{\text{speed}} - \bar{V}$ using data of 970 kilometers roads and 1000 drivers. Accident rate is low when vehicles are traveling at a speed within one standard deviation around the average speed. When speed reaches 10 km/h more

than the average speed (V_{85}), accident rate reaches minimum. U-curve is suitable for steady flow and it is the theoretical basis of Δv_{85} limit method. However, a single speed limit is not accorded with the normal speed trend. In addition, accident data requires long-term accumulation. It means that accident data is not easy to obtain.

- (3) Driving behavior is complex: When driver approaches the exit ramp from the main line, they need to finish a series of complex driving behaviors, such as looking for acceptance gap, slowing down, and change lanes in the slow lane. Driver's driving habits and reflections are different in the different area or on different roads [14]. Since different researchers developed different models, it is necessary to know whether the value of the speed limit designed for a particular freeway can be applied to another freeway in the same area and whether the value of the speed limit determined for one specific road in specific areas could imply in another road in another area.
- (4) Consider geometry parameters only: The comprehensive influential factors method only considers geometric parameters, such as the slope and curve radius. McLean [15] studied the level of the expected speed and radius regression curve according to the data from Australia and New Zealand. The conclusion shows that when the curve radius is greater than 1000 m, 85% of the expectation speed is not affected by geometry parameters. In China, JTG/T B05-20049 (Guidelines for Safety project on Highway) [16] proposed that the sections whose curve radius is less than 1000 m and the absolute value of slope is less than 3% can be considered as a straight-line. Above studies illustrated that in some sections, geometry parameters should not be the only parameter in consideration.

3. Research Goal

In this paper, the primary goal is to develop a successive-stage speed limit model for vehicle speed along the exit upstream ramp of direct-type freeway in China. The specific tasks of this paper can be summarized as follows:

- (1) State the necessity of operating successive-stage speed limit control on the exit ramp upstream.
- (2) Analyze the relationship between operation speed and deceleration lane length on the exit ramp upstream, and prove that the speed limit should be in accordance with the deceleration lane length.
- (3) Determine whether the speed limit needs to be set for the small vehicle and large vehicles separately on the exit ramp upstream.
- (4) Analyze the main factor of the speed limit from saturation, turning ratio, and vehicle type using double-factors curve fitting and polynomial regression.
- (5) Build speed-prediction model according to the result of factor analysis.
- (6) Calibrate speed-limit sign validity model by linearization
- (7) Establish successive-stage speed-limit model based on a speed-prediction model and a speed-limit sign validity model.
- (8) Verify the validity of the successive-stage speed-limit model by a case study.

4. Hypothesis and Research Objective

4.1. Hypothesis

The research is based on three hypotheses.

- (1) The speed limit for the exit ramp nose is reasonable. This research focuses on developing a successive-stage speed limit model on the exit ramp upstream. The value of the speed limit of the exit ramp nose is not studied in this paper; therefore assume that the speed limit on the exit ramp nose is reasonable.
- (2) 90% of right-turn drivers have finished lane change at 2/3 location of deceleration lane in China [17]. To simplify driver behavior, all vehicles are assumed to have finished lane change.
- (3) Except mountainous freeway, geometric line of freeway meets freeway design standard. Combining with the actual situation, as well as the result of McLean [15] study, and JTG/T B05-2004 [16], the line of the exit ramp upstream recognized as the line.

4.2. Research Objective

The type of an exit ramp has influences on the traffic flow of the exit ramp, speed distribution and driver behavior. So the exit ramp type should be determined firstly. 428 exit ramps from 11 provinces in China were observed by Google Earth. Statistically, 93.5% of the ramps fell into the above 4 categories by the change of lanes number in the main line of upstream and downstream, the setting of the deceleration lane, the number of ramp lane, and the separation situation of ramp. 55.8% of them were direct-type exit ramp. Hence, direct-type exit ramp is the research objective in this study. Characteristics of direct-type exit ramp are as follows:

- (1) The number of lanes in the main line of upstream and downstream is constant.
- (2) The area of ramp upstream is broadened properly and has a deceleration lane.
- (3) The number of lane on the ramp is 1.
- (4) The exit ramp lanes are not separated.

5. Data Collection and Process

5.1. Site Selection

Sites are selected as follows:

- (1) Safeguard facilities at the road side are in normal condition.
- (2) Has a good visual space.
- (3) Service level is A or B (to ensure that speed is not significantly affected by other vehicles in the flow).
- (4) The freeways selected in this study should include two lanes, three lanes, and four lanes in one direction in the main line.
- (5) The range of deceleration lane length is 150–250 m.

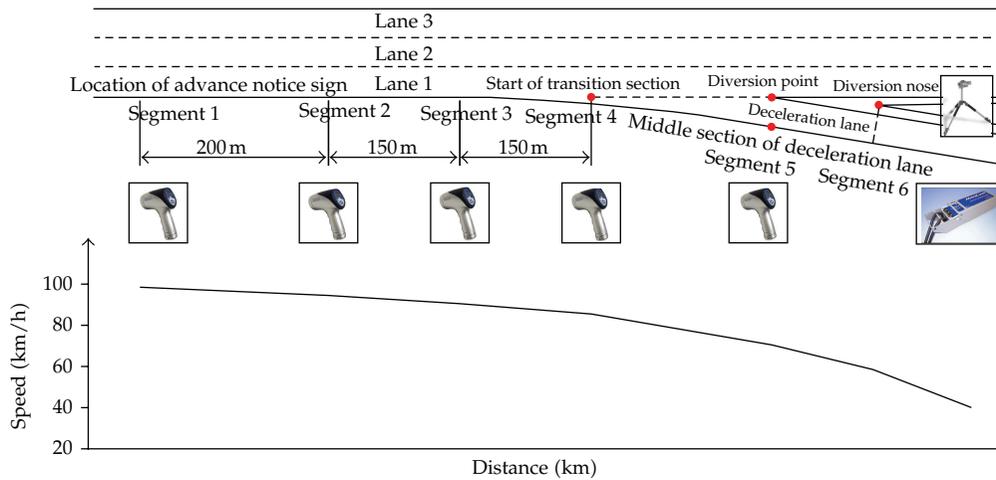


Figure 1: Speed profile of the exiting flow.

- (6) The values of the speed limit on the main line are 100 km/h and 120 km/h; the values of the speed limit on exit ramp are 40 km/h and 60 km/h, respectively.
- (7) No congested traffic phenomenon.

As a result, seven exit ramps in Nanjing, China were chosen as candidate sites.

5.2. Investigation Segments

The exit ramp impact area is within 500 meters upstream of nose, according to HCM, as length 500 m from the nose [18]. So the research area is from the location of the advance notice sign to nose and is divided into six segments according to deceleration trend. As shown in Figure 1, Segment 1 is the location of the advance notice sign; segment 4 is the start of the transition section; segment 5 is the middle section of the deceleration lane; Segment 6 is cross-sectional on the nose. Data collection area is the upstream of main road lanes. The curve shows the speed-change trend in this section in Figure 1.

5.3. Data Collection

Data in this study includes geometric parameters, traffic-flow parameters and traffic-control parameters.

Geometric parameters include alignment elements and lane number in the main lanes and deceleration lane length. It is known that the alignment is not considered in above studies. Therefore, geometric parameters refer to the lanes number in the main lanes and deceleration lane length. The main traffic control parameter is main lane speed limit and exit ramp speed limit. Deceleration behavior on the main line mainly happens on the outside lane. Traffic-flow parameter includes traffic volume and point speed of the outside lane.

Traffic flow parameters are collected during the peak hour from 8 to 11 a.m or 2 to 5 p.m on a clear, well visibility, and regular-temperature day.

The procedures of data collection are designed as follows:

- (1) *Preparation*: Train 8 investigators; prepare 2 radar guns (LaiSai Bushnell), 1 MetroCount 5600 and 1 camera (Sony Handcam). Investigators should be familiar with the investigation plan, the survey locations, and can use the guns, MetroCount 5600, and camera. Radar gun and MC5600 are applied to collect point speed and camera is applied to collect traffic volume.
- (2) *Arrangement equipments*: Put 4 radar guns, MC5600, and Sony Handcam camera at the location of Sections 1 to 6 respectively, as shown in Figure 1.
- (3) *Observer assignment*: Assign two observers at each location of Sections 1, 2, 3, and 4, respectively. One is the radar-gun holder and the other one is the recorder.

The radar-gun holder is in charge of shooting speed, reading the last three digits of the plate number, and informing the recorder of the information.

Observers are hidden from traffic to minimize the effect of their presence on passing vehicles.

5.4. Data Process

The data collected from the camera were sorted into small-vehicle and large-vehicle categories. The small vehicle has less than 20 seats for passenger vehicles or less than 2 tons in weights for freight vehicles.

It was found that Δv_{85} is a major parameter used by traffic engineers and transport planners from literature. Point speed will be transformed to Δv_{85} .

Discard nonnormal data that does not conform to normal distribution using statistic software SPSS [19].

6. Data Analysis

6.1. Δv_{85} Characteristics Analysis

Fitzpatrick [20] divided Δv_{85} into 3 groups: $\Delta v_{85} \leq 10$ km/h, 10 km/h $< \Delta v_{85} < 20$ km/h, and $\Delta v_{85} \geq 20$ km/h.

- (1) $\Delta v_{85} \leq 10$ km/h means that operation speed is well and accident rate is low.
- (2) 10 km/h $< \Delta v_{85} < 20$ km/h means that the coordinated of operation speed becomes bad, and the accident rate has increased.
- (3) $\Delta v_{85} \geq 20$ km/h, means that the coordinated operation speed becomes worse, and the accident rate is high.

It is worth noticing that the accident rate of observations with $\Delta v_{85} \geq 20$ km/h is 6 times more than that of $\Delta v_{85} \leq 10$ km/h. We can see that the higher the Δv_{85} is, the more the accident rate is. Hence, Δv_{85} should be lower than 10 km/h. If $\Delta v_{85} \geq 10$ km/h, successive-stage speed limit should be applied. Figure 2 shows the Δv_{85} curves of the exit ramp upstream.

In Figure 2, the Δv_{85} of all sites are less than 10 km/h in segments 1–4. It reveals that the trend of speed is gentle and the accident rate is low in segments 1–4. In segments 4–6, the Δv_{85} of Sites 1, 2, 3, and 7 are more than 10 km/h. Especially, the Δv_{85} of Site 2 is close

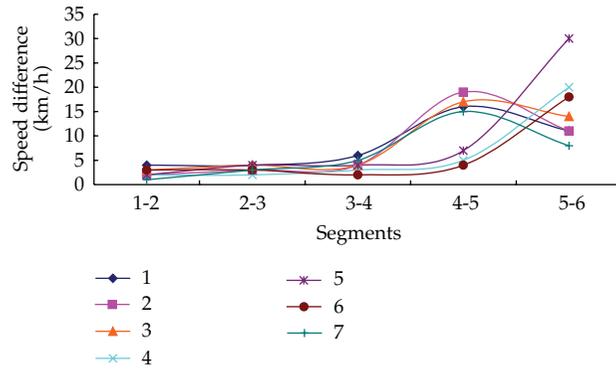


Figure 2: Δv_{85} curve on the exit ramp upstream.

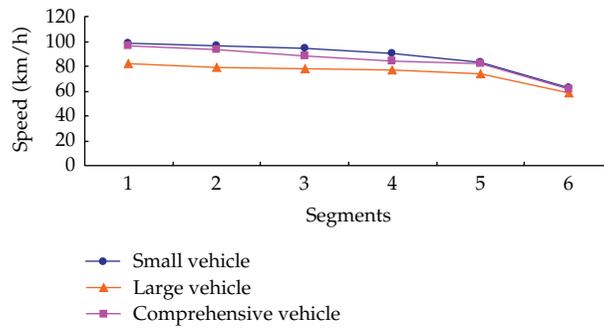


Figure 3: v_{85} curve of small, large, and comprehensive vehicles.

to 20 km/h. It shows that the accident rate has increased in segments 4–6. Δv_{85} of Sites 4, 5, and 6 are higher than 20 km/h as in segments 5–6. In this scope, the accident rate increases significantly. Δv_{85} of Sites 1, 2, 3, and 7 increase from 5 km/h to 15 km/h approximately from segments 3–4, then decreases from 15 km/h to 10 km/h approximately. For Sites 4, 5, and 6, Δv_{85} dramatically increases from 5 km/h to 20 km/h approximately from segments 4–6.

Through the above analysis, we can conclude that it is necessary to operate successive-stage-speed limit control on the exit ramp upstream.

6.2. The Speed of Vehicle Type Characteristics Analysis

The v_{85} curves of small, large and comprehensive vehicles are shown in Figure 3.

There are three-speed curves in Figure 3. The above curve is small vehicles; the middle curve is comprehensive vehicles; and the below curve is large vehicles. Comprehensive vehicle represents the speed level for the small vehicle and the large vehicle. In Figure 3, curves of small vehicles and large vehicles have a similar trend. Specific performance is that they all decrease slowly in segments 1–4 and a rapid decline in segments 4–6. Meanwhile, difference of v_{85} does not exceed 10 km/h in segments 1–4 and 5 km/h in segments 4–6. The curve of the comprehensive vehicle is between the small vehicle curve and large vehicle curve. In segments 4–6, v_{85} curves of the small vehicle and large vehicle approach v_{85}

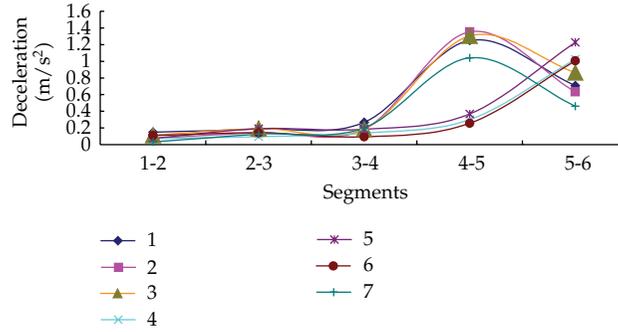


Figure 4: v_{85} speed-deceleration-lane-length curve.

curves of comprehensive vehicle simultaneously. Finally, three curves intersect on 65 km/h at segments 7. Hence, speed limits of small vehicles and large vehicles are same.

It is not needed to design different speed limits for small vehicles and large vehicles, respectively, on the exit ramp upstream, as the speed limit control pattern on the main line.

6.3. The Deceleration Lane Length Characteristics Analysis

According to hypothesis 3, the only difference of the seven sites in terms of geometric aspect is the deceleration lane length. Speed-deceleration-lane-length curve is used to judge whether the speed limit needs to be determined according to deceleration lane length.

The range of deceleration lane length used in this paper is 150–250 m. The range almost covers all deceleration-lane-length of freeway in China. If there are different trends in speed-deceleration lane length curves in Sites 1, 2, 3, 4, 5, 6, and 7, the speed limit should be determined due to the differences of the deceleration lane length.

Figure 4 is the speed-deceleration-lane-length curve.

In segments 1–4, curves of all 7 sites are nearly flat. In segments 4–6, however, there are two opposite trends. The curves of Sites 1, 2, 3, and 7 rise in segments 4–5, and then drop in segments 5–6. The trends of the curve 4, 5, 6 are opposite. The common characteristics of 1, 2, 3, and 7 is that their deceleration lane length are more than 200 m. The lengths of the deceleration lane of 4, 5, and 6 are less than 200 m.

Hence, deceleration lane length is related to v_{85} and the speed limit should be different according to the deceleration lane length.

6.4. The Speed of Lanes Number of Main Line Characteristics Analysis

In order to analyze whether the number of lanes of main line affect speed, speed-lane number curve is displayed in this section. The numbers of lanes in the main line for one direction are 2, 3, 4 in this study.

Speed-lane number curves are shown in Figure 5.

In segment 1 to segment 4, the curves are similar to each other. The difference of the curve is manifested in 4–6. The highest speed level appears in four lanes, and then in three lanes and two lanes. The trend of the curve is similar with the curve in Figure 4. It can be explained that the difference of speed-lane number curve is caused by different deceleration lane lengths.

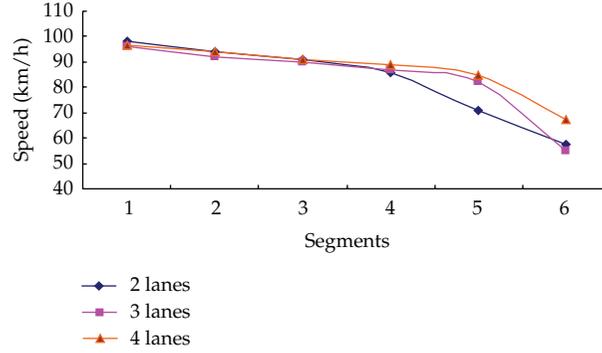


Figure 5: Speed-lane number 10 curves.

We can deduce that the lane number in the main line does not relate to speed, and lane number in the main line has not been considered when we determine the speed limit.

$\Delta v_{85} \geq 10$ illustrates the necessity of speed limit on the exit ramp. Speed-deceleration-lane-length curve presents two trends bounded by 200 m, so the speed limit should be in accordance with the deceleration length.

7. Successive-Stage Speed Limit Model

7.1. The Frequency of the Speed Limit

Many drivers become irritated by frequency of slow speed. Considering the maneuverability of the speed limit sign, the number of successive-stage speed limit should not be more than two stages.

7.2. Model Choice

The speed limit has a certain impact on operational speed, but not all vehicles travel under the speed limit. In 1986, Anders and Arne [21] analyzed the relationship between initial speed under different speed limits and drivers' behavior. A speed limit influential model is developed under free flow circumstance and shown as follows:

$$V_T = \frac{V_0}{1 + c \cdot d^{z^2}}, \quad (7.1)$$

$$z = \frac{V_g}{V_0},$$

where V_T is the operational speed limit area (km/h), V_0 is the operational speed in the upstream of the speed limit area (km/h), V_g is the speed limit (km/h), c , d , are influential coefficients.

Equation (7.2) is the linearization of (7.1)

$$\frac{V_0}{V_T} - 1 = c \times d^{z^2}, \quad (7.2)$$

$$\ln\left(\frac{V_0}{V_T} - 1\right) = \ln c + z^2 \ln d.$$

Set $Y = \ln(V_0/V_T - 1)$, $a = \ln c$, $b = \ln d$, $X = z^2$, it can be obtained that

$$Y = a + bX. \quad (7.3)$$

It is can be found that R Square (goodness of fit) is 0.839, Adjusted R Square (Coefficient of Determination) is 0.838, and Sig. (significant indicators) are less than 0.05. The results meet the statistical requirements. So that can be obtained that

$$Y = 0.903 - 3.657X, \quad (7.4)$$

$$V_T = \frac{V_0}{1 + 2.467 \times 0.026z^2}.$$

V_0 can be obtained by field investigation. For new freeway, V_0 can only be obtained through speed prediction model. The building process of speed-prediction model is studied in the next section.

7.3. Speed Prediction Model

7.3.1. Influence Factors Analysis

It is widely believed that influence factors of speed include geometric parameters, traffic-flow parameters, and traffic control parameters. Traffic-flow parameters include traffic volume, speed, density, average time headway, average space headway, turning ratio and vehicle type, and so forth. Traffic flow parameters comprise density, average time headway, and average space headway. Therefore, traffic volume, turning ratio and vehicle type are chosen in this paper. To increase the universality of the model, traffic volume is converted to saturation.

7.3.2. Speed Prediction Model

Double-factor curve fitting is used to analyze the correlation between various factors and speed by *SPSS*. From test results, the R^2 value and P value of the quadratic curve and cubic curve based on saturation and large vehicles rate meet statistical requirement and these are used as independent variables. But models of turning ratio and operational speed do not meet statistical requirement, so turning ratio is removed. The form of the quadratic curve is simple, so quadratic curve is selected as the basic model in this paper.

A two-factor multinomial model is developed to compare how saturation and percentage of large vehicles affects operational speed. Independent variables are selected by the stepwise regression method. Every variable selected by the stepwise regression method is tested. Percentage of large vehicles is eliminated in *SPSS*.

Saturation is selected as the main factor affecting operational speed. Step length needs to be decreased to increase model fit. The Adjusted region of saturation is p [0.3, 0.7] with a step length of 0.5 and 9 levels. Stepwise regression is analyzed in *SPSS*. Variation of saturation is eliminated in regression, and the calibrated model is as follows:

$$V = 104.788 - 13.465S^2, \quad (7.5)$$

where S is saturation.

Table 1: Average deceleration in each section.

Deceleration lane length (m)	a_t (m/s ²)	a_{02} (m/s ²)	a_{01} (m/s ²)
<200 m	0.15	1.30	0.73
>200 m		0.31	1.09

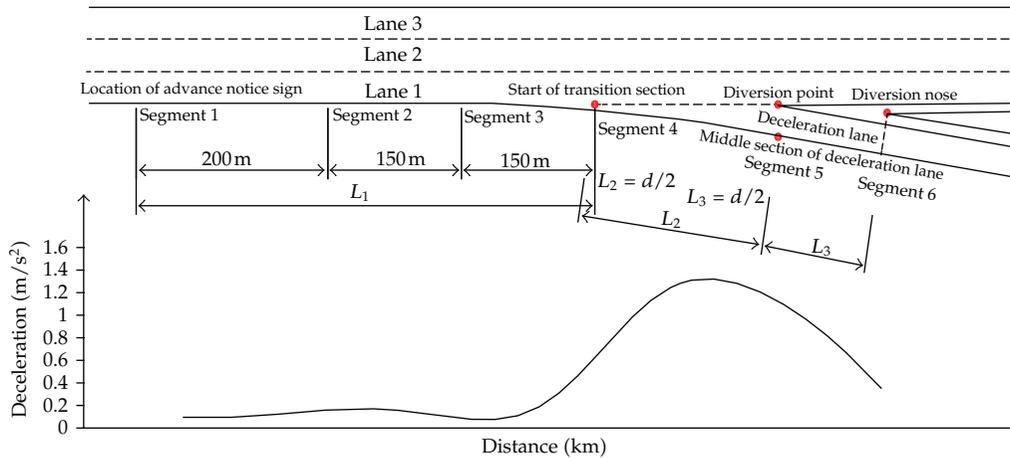


Figure 6: Segments of research scope divided by deceleration.

Hypothesis tests are carried out in SPSS. Outputs are R square is 0.844, Sig. is 0.000, these manifest that the goodness-of-fit and coefficients test is statically satisfied. The residuals and dependent variable are distributed as normal distribution approximately. All of the results prove that the speed-prediction model proposed in this paper is statistically significant.

7.4. Deceleration Value

It can be concluded that the variation of deceleration is more significant than speed difference and speed. The research scope is divided into three sections according to the variation of the deceleration. The entire exiting process is divided into three sections below, as shown in Figure 6. The curve shows the deceleration change trend in this section in the following Figure 6.

In Figure 6, speed difference in L_1 section is low and barely changes. Suppose that vehicles in L_1 travel by the average deceleration are denoted as a_t from Sections 1 to 4. Decelerations on L_2 and L_3 are denoted as a_{01} and a_{02} which are average field decelerations. Average deceleration in each section is illustrated in Table 1.

7.5. The Speed Limit Value

Denote V'_t as the reasonable speed of the middle section of the deceleration lane and V_{go} as the speed limit at the exit ramp painted nose. During constant deceleration (a_{01}), the speed of

vehicles changes from V_{go} to V'_t in L_3 . V'_t can be derived according to the kinematics principle, and the expression is

$$V'_t = 3.6\sqrt{a_{01}d + \left(\frac{V_{go}}{3.6}\right)^2}, \quad (7.6)$$

where V'_t is the reasonable speed at the middle section of deceleration lane (km/h); V_{go} is the speed limit at the exit ramp painted nose (km/h); a_{01} is the average deceleration at L_3 in the exit ramp area (m/s^2).

In L_2 , it is known that vehicles move with a constant acceleration of a_{02} . Define V_t as a reasonable speed of the start of the transition section. It can be expressed as

$$V_t = 3.6 \times \sqrt{a_{02}d + \left(\frac{V'_t}{3.6}\right)^2}, \quad (7.7)$$

where V_t is the reasonable speed of the start of transition section (km/h); a_{02} is average deceleration at L_2 in the exit ramp area (m/s^2).

Combine (7.6) and (7.7), V_t can be calculated using a_{01} and a_{02} .

$$V_t = 3.6 \times \sqrt{(a_{01} + a_{02})d + \left(\frac{V_{go}}{3.6}\right)^2}. \quad (7.8)$$

At L_1 , vehicles move with a constant acceleration of a_t . Define V_T as reasonable speed at the start of the outside lane of deceleration. During constant deceleration a_{01} , the speed of vehicles changes from V_T to V_t . V_T can be expressed as

$$V_T = 3.6\sqrt{2a_tL_1 + \left(\frac{V_t}{3.6}\right)^2} = 3.6\sqrt{2a_tL_1 + (a_{01} + a_{02})d + \left(\frac{V_{go}}{3.6}\right)^2}, \quad (7.9)$$

where, V_T is reasonable speed at the start of the outside lane of the deceleration (km/h); a_t is average deceleration at L_1 in the exit ramp area (m/s^2); L_1 is the length of the exit ramp upstream section (500 m).

Define V'_g as the successive-stage speed limit at the outside lane in the main line. The successive-stage speed limit was deduced based on V_0 of a field investigation or of the speed-saturation prediction model and V_T of speed-limit sign validity model.

$$V'_g = V_0\sqrt{\frac{0.903 - \ln(V_0/V_t - 1)}{3.650}}, \quad (7.10)$$

where, successive-stage speed limit at the outside lane of main line (km/h); V_0 is operational speed at the outside lane in the exit ramp upstream gained though a field investigation or speed-saturation prediction model.

7.6. The Placement of Speed-Limit Sign

At present, the placement of the speed-limit sign is determined by the basis of the psychical process of drivers perceiving and reacting to signs. The determinant of the perception-reaction process is the placement of the danger point. Thus, the paper is focused on a danger point to confirm the placement of successive-stage speed limit in this section.

By analyzing deceleration characteristics curves, we found drivers began to slow down in segment 1. Between segment 1 and segment 2, curves decreased smoothly and steadily. A significant change in deceleration happened in segment 2–4. According to *Investigation segments* section, we can find that advance notice sign is located at segment 1. It can explain the trend of deceleration in segment 1–4 commendably. So the location of segment 1 is identified as a danger point.

The placement and clear height of successive-stage speed limit sign are determined according to *Traffic Engineering Manual* [22].

8. Case Study

Yang Dongfang exit ramp along Nanjing Round Freeway is chosen to verify the feasibility and validity of the successive-stage speed limit model.

After investigation, we got the information about *Yang Dongfang* exit ramp. The length of the deceleration lane is 170 m. From Table 1, this paper takes the values of a_{01} , a_{02} and a_t to be 0.73 m/s^2 , 1.3 m/s^2 , and 0.15 m/s^2 , respectively. Vehicles speed in the main line is controlled below 120 km/h and exit line is controlled under 40 km/h. So, $V_{g0} = 40 \text{ km/h}$. Operational speed in the outside lane on the exit ramp upstream is 98 km/h. So, $V_0 = 98 \text{ km/h}$.

Using successive-stage speed limit model, the value of successive-stage speed limit can be gained.

$$\begin{aligned}
 V'_t &= 3.6 \times \sqrt{a_{01}d + \left(\frac{V_{g0}}{3.6}\right)^2} = 3.6 \times \sqrt{0.73 \times 170 + \left(\frac{40}{3.6}\right)^2} = 56.6 \text{ km/h}, \\
 V_t &= 3.6 \times \sqrt{a_{02}d + \left(\frac{V'_t}{3.6}\right)^2} = 3.6 \times \sqrt{1.3 \times 170 + \left(\frac{56.6}{3.6}\right)^2} = 77.9 \text{ km/h}, \\
 V_T &= 3.6 \times \sqrt{2a_tL_1 + \left(\frac{V_t}{3.6}\right)^2} = 3.6 \times \sqrt{2 \times 0.15 \times 500 + \left(\frac{77.9}{3.6}\right)^2} = 89.5 \text{ km/h}, \\
 V'_g &= V_0 \sqrt{\frac{0.903 - \ln(V_0/V_T - 1)}{3.650}} = 98 \times \sqrt{\frac{0.903 - \ln(98/89.5 - 1)}{3.650}} = 92.6 \text{ km/h}.
 \end{aligned} \tag{8.1}$$

Combined with engineering experience, get the value of successive-stage speed limit is 90 km/h. This result suggests that successive-stage speed limit is not determined for the small vehicle and large vehicle, respectively, on the exit ramp upstream.

9. Discussion and Conclusions

v_{85} is a major parameter used by traffic engineers and transport planners. Generally, speed limits are set at or below the Δv_{85} [23]. A bill to raise Texas highway speed limit to 85 mph could have motorists getting there faster but shelling out much more money at the gas pump [24]. The paper uses the “85th percentile speed” to determine the successive-stage speed limit.

The previous studies show when the curve radius is greater than 1000 m, 85% of the desired speed is not affected by horizontal alignment. JTG/T B05-2004 [16] proposed that the sections which the curve radius are less than 1000 m and slope of absolute value small than 3% may be handled as a straight-line section. Except the mountainous freeway, geometric alignments of most freeways meet freeway design standard. Above studies illustrated that geometry parameters should not be the only parameter in consideration.

In segments 4–6, the Δv_{85s} of Sites 1, 2, 3, and 7 are more than 10 km/h. Especially, the Δv_{85} of Site 2 is close to 20 km/h. Δv_{85s} of Sites 4, 5, and 6 is more than 20 km/h in segments 5–6. These show that accident rate has increased in segments 4–6. Through the above analysis, it is necessary to set successive-stage speed limit control on the exit ramp upstream.

Curves of the small vehicle and large vehicle have a similar trend. Meanwhile, difference of v_{85} does not exceed 10 km/h in segments 1–4 and 5 km/h in segments 4–6. The curve of the comprehensive vehicle is between curves of the small vehicle and the large vehicle. In segments 4–6, v_{85} curves of the small vehicle and large vehicle approach the curve of comprehensive vehicle simultaneously. Finally, three curves intersect on 65 km/h at segment 7. Hence, speed limits of the small vehicle and large vehicle are same. Case study also tests this conclusion. So, successive-stage speed limit is needed to be developed for the small vehicle and large vehicle, respectively, on the exit ramp upstream.

In segments 4–6, deceleration curves show two opposite trends. The curve of Sites 1, 2, 3, and 7 rise in segments 4–5, and then drop in segments 5–6. The trend of the curve of 4, 5, and 6 is opposite. The common characteristic of 1, 2, 3, and 7 is that their deceleration lane length is more than 200 m. The length of the deceleration lane about 4, 5, and 6 is less than 200 m. Hence, deceleration lane length is related with v_{85} and the speed limit should be in accordance with the deceleration lane length. This means deceleration lane length is an important factor for successive-stage speed limit.

Speed-lane number curves in the main line are similar with deceleration curves. The similarity suggests that the difference of speed-lane number curves is caused by different deceleration lane length. We can deduce that lane number in the main line does not relate with speed and the lane number of the main line has not been considered when determine the speed limit.

Double-factor curve fitting is used to analyze the correlation between saturation, turning ratio, and large vehicles rate and speed. Based on test results of models, saturation is selected as the main factor of operational speed. As a result, the paper built speed-saturation prediction model.

The research scope is divided into three sections according to the variation of the deceleration. Based on field investigation speed or speed-saturation prediction model and the speed-limit sign validity model, the value of successive-stage speed limit is deduced according to kinematic principles.

The conclusions are summarized as follows:

- (1) It is necessary to set successive-stage speed limit control on the exit ramp upstream. The successive-stage speed limit is not needed to be developed for small vehicles and large vehicles, respectively, on the exit ramp upstream.
- (2) Geometry parameters are not considered in building successive-stage speed limit model.
- (3) Deceleration lane length is related to v_{85} , and the speed limit should be in accordance with the deceleration lane length.
- (4) Lane number in the main line does not relate to speed and cannot be considered when determining the speed limit.
- (5) Saturation is the main factor affecting speed, and the paper built a speed-saturation prediction model.

Acknowledgments

This work was supported by the National Natural Science Foundation (51078232), Jiangsu Provincial Communications Department Science Foundation (2010Y17), Hebei Provincial Communications Department Science Foundation (Y-2010016).

References

- [1] M. Firestine, H. McGee, and P. Toeg, "Improving truck safety at interchanges: final report to the Federal Highway Administration," FHWA IP-89-024, U.S. Department of Transportation, Washington, DC, USA, 1989.
- [2] C. N. Kloeden, A. J. McLean, and G. Glonek, "Reanalysis of travelling speed and the risk of crash involvement in Adelaide South Australia," Report CR207, Australian Transport Safety Bureau, Canberra, Australia.
- [3] M. Peden, R. Scurfield, D. Sleet et al., *World Report on Road Traffic Injury Prevention*, World Health Organization, Geneva, Switzerland, 2004.
- [4] G. Milliken, "Managing speed: review of current practice for setting and enforcing speed limits," *Transportation Research Board*, pp. 36–70, 1998.
- [5] W. C. Taylor, *Speed Zoning Guidelines: A Proposed Recommended Practice*, Institute of Transportation Engineers, Washington, DC, USA, 1990.
- [6] G. Milliken, "Managing speed: review of current practice for setting and enforcing speed limits," Special Report 254 [R], Transportation Research Board, Washington, DC, USA, 1998.
- [7] K. R. Agent, J. G. Pigman, and J. M. Weber, "Evaluation of speed limits in Kentucky," *Transportation Research Record*, no. 1640, pp. 57–64, 1998.
- [8] A. A. M. Aljanahi, A. H. Rhodes, and A. V. Metcalfe, "Speed, speed limits and road traffic accidents under free flow conditions," *Accident Analysis and Prevention*, vol. 31, no. 1-2, pp. 161–168, 1999.
- [9] J. Park, *Modeling of setting speed limits on urban and suburban roadways [Ph.D. thesis]*, University of South Florida, 2003.
- [10] NCHRP, "Safety impacts and other implications of raised speed limits on high speed roads," Contractors Final Report, 2006.
- [11] DDS, *2011 Driver's Manual*, The Department of Driver Services, 2011, <http://www.dds.ga.gov/docs/forms/fulldriversmanual.pdf>.
- [12] A. Edgar and M. Tziotis, "Computerizing Road Safety," in *Australia Road Research Board*, pp. 46–51, 1999.
- [13] D. R. Solomon, *Accidents on Main Rural Highways Related To Speed, Driver and Vehicle*, US Department of Commerce, Federal Bureau of Highways, Washington, DC, USA, 1964.
- [14] W. H. Wang, W. Zhang, H. W. Guo et al., "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research Part C-Emerging Technologies*, vol. 19, no. 6, pp. 1202–1214, 2011.

- [15] J. McLean, "Driver speed behavior and rural road alignment design," *Traffic Engineering & Control Magazine*, pp. 208–211, 1981.
- [16] CHELBI Engineering Consultants. INC., *Guidelines for Safety Project on Highway (JTG/T B05-2004)*, Communications Press, 2004.
- [17] R. H. Rong, *Research on Traffic Characteristic of Interchange Diverging Area [D]*, Southeast University, 2006.
- [18] National Research Council Washington, D.C. *Highway Capacity Manual 2000*, Transportation Research Board, 2000.
- [19] C. Henry Jr. and Thode, *Testing For Normality*, Marcel Dekker, New York, NY, USA, 2002.
- [20] K. Fitzpatrick, "Evaluation of design consistency methods for two-lane rural highways," Publication FHWA-RD-99-174. FHWA, U.S. Department of Transportation, 1999.
- [21] B. Anders and C. Arne, *The VTI Traffic Simulation Model*, Swedish Road and Traffic Research Institute, 1986.
- [22] Transportation Research Board, *Traffic Engineering Manual*, Department of Transportation, 2000.
- [23] Speed Zoning Information, "Institute of transportation engineers[DB/OL]," 2004, http://www.ite.org/standards/speed_zoning.pdf.
- [24] Lyneka, "Texas may raise speed limit to 85 MPH," <http://abcnews.go.com/Business/texas-moves-raise-speed-limit-85-mph/story?id=13319173>.

Research Article

A Model for Bus Crew Scheduling Problem with Multiple Duty Types

Mingming Chen 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 18 June 2012; Revised 23 August 2012; Accepted 26 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 M. Chen 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.

This paper presents an approach for solving the bus crew scheduling problem which considers early, day, and late duty modes with time shift and work intensity constraints. Furthermore, the constraint with the least crew number of a certain duty (e.g., day duty) has also been considered. An optimization model is formulated as a 0-1 integer programming problem to improve the efficiency of crew scheduling at the minimum expense of total idle time of crew for a circle bus line. Correspondingly, a heuristic algorithm utilizing the tabu search algorithm has been proposed to solve the model. Finally, the proposed model and algorithm are successfully tested by a case study.

1. Introduction

The urban bus crew scheduling problem is a fundamental and crucial problem arising in the field of transit scheduling, because there are the main resources necessary to service passengers. Given the timetable of an urban transit line during one day, the scheduling problem of a bus crew is to complete the operating arrangement of all crews with different duty types. The duty types of crew can be divided into three types: early, day, and late modes. With the influence of working hours, place of residence, living habits, and other factors, there are different requirements to the crew. For example, the crew can only choose trips with day duty when living farther away from the bus depot, the crew can select trips with early or late duty when they have temporary activities at noon, the crew can choose trips with early or day duty when they need to take care of family members at night, and some crews can be allocated to trips with any kind of duty type. Therefore, how to arrange the crew scheduling so as to utilize the manpower resources effectively and satisfy a set of constraints laid down by the labor regulations, has an important theoretical and practical significance.

The urban bus crew scheduling has received much attention in the last decade, and there have been many models and solution methods. According to the objective of model,

it can usually be divided into set covering, set partitioning, and multiobjective program. Smith and Wren [1] introduced a crew scheduling problem formulation based on set covering problem, solved by using integer linear programming. The elements covered were duty pieces, and the covering process used slack and surplus variables. Darby-Dowman and Mitra [2] presented a natural generalization called the extended set partitioning model which formed the basis of a computer assisted bus crew scheduling system. An integrated vehicle and crew scheduling problem proposed by Mesquita and Paias [3] were described by an integer linear programming formulation combining a multicommodity network flow model with a set partitioning/covering model. Beasley and Cao [4] developed a new lower bound for the crew scheduling problem based on dynamic programming. Lourenco et al. [5] brought a multiobjective crew scheduling model, which was tackled using a tabu search technique, metaheuristics, and genetic algorithms. Huisman et al. [6] presented an integrated approach to solve a vehicle scheduling problem and a crew scheduling problem on a single bus route.

As for the solution method of bus crew scheduling, it can be divided into mathematical programming methods including integer linear programming, column generation, branch and bound method, and heuristic methods including genetic algorithm, tabu search algorithm, colony algorithm, and so forth.

Mitra and Darby-Dowman [7] established a computer-based bus crew scheduling model using integer programming. The approach presented by Desrochers and Soumis [8] used column generation solution procedure. Clement and Wren [9] introduced a solution for the crew scheduling problem using a genetic algorithm, and several greedy algorithms were used for assigning duties to pieces of work. Beasley and Cao [10] established a 0-1 integer programming model, and a tree-search algorithm was used to obtain the final optimum. Ceder [11] used deficit function properties to construct vehicle chains, splitted and recombined vehicle blocks into legal duties, and used a shortest-path and matching algorithm for solving the problem. Shen and Ni [12] established a crew scheduling model with minimizing the number and cost of duty and designed the multineighbourhood structure. The crew scheduling model was abstracted by Yang et al. [13] as an optimization problem for the minimum of a nonlinear constraint function, and a genetic algorithm was used to solve the problem. Niu [14] focused on how to determine the skip-stop scheduling for a congested urban transit line during the morning rush hours and formulated a nonlinear programming model with the objective of minimizing the overall waiting times and the in-vehicle crowded costs, and a bilevel genetic algorithm was proposed to solve the problem. Srdjan et al. [15] proposed architecture for a new scheduling mechanism that dynamically executes a scheduling algorithm using feedback about the current status grid nodes, and two artificial neural networks were created in order to solve the scheduling problem. A tabu search algorithm was proposed by Atli [16] to provide good solutions to resource-constrained, deterministic activity duration project scheduling problems.

For an overview on these papers, although a lot of attention has been paid to the factors associated with the required crew number, total work hours, layover and rest time, impartiality, and so forth in bus crew scheduling, only a few papers consider the duty type constraint, which should be well considered in process of urban transit operation.

The remainder of paper is organized as follows. In Section 2, we describe the urban bus crew scheduling problem and establish an optimization model with minimizing the total crew idle time. A tabu search algorithm is then designed in Section 3. In Section 4, a numerical example is provided to illustrate the application of the model and algorithm. Finally, we present the conclusions and some comments on topics requiring future research.

2. Urban Bus Crew Scheduling Model

2.1. Problem Description

Given a set of trips within a fixed planning horizon, the crew scheduling problem consists of generating a feasible set of crew schedule which includes sign-on and sign-off time, crew duty types, and a serial of trips which covers the bus schedule and satisfies a range of constraints laid down by work rules. The trips that are assigned to the same crew member define a crew duty. Together the duties constitute a crew schedule. The criteria for crew scheduling are based on an efficient use of manpower resources while maintaining the integrity of any work rule agreements. The usual crew scheduling problem uses one or more of these objectives: minimum sum of crew costs, minimum number of crew required, and maximum the efficiency of crew usage.

The crew scheduling problem in this study is to arrange the work plan for crew in a single bus line. The kind of urban bus line is a circle one, which means that the starting station and terminal are the same bus station. Each trip has fixed departure and arrival times and can be assigned to a crew member from a determined depot. All trips have been numbered in ascending departure time order. Each crew should be restricted to work in the same vehicle during the bus operating time.

In particular, each crew duty must satisfy several complicating constraints corresponding to work rule regulation. Typical examples of such constraints are maximum working time without a break, minimum break duration, maximum total working time, and maximum duration. In this paper, the considered work rules mainly involve time shift, work intensity, and duty type compatibility constraints. The time shift constraint means that the difference between departure time of latter trip and arrival time of former trip at the same station should not be less than the minimum layover time, as for the two adjacent trips carried by the same crew. The work intensity constraint can be defined that the total work time of each crew does not exceed the maximum number of work hours corresponding to the duty type of crew.

According to the different peak hours, crew duty can be divided into three types, namely, early, day, and late mode. The early duty covers morning peak hours and the range of working time is generally from 6:00 to 13:00, while the late duty covers evening peak hours, and the range of working time is generally from 15:00 to 22:00. Due to the evening peak hours forward and morning peak hours backward in weekend, the day duty covers two peak hours, and the range of working time is from 9:00 to 18:00. In this paper, we consider the three duty types mentioned before, and the sign-on and sign-off time can be chosen flexibly with work plan. The duty type compatibility constraint means that the departure time of trip should be later than the earliest sign-on time and earlier than the latest sign-off time corresponding to crew duty type.

With the different working time of duty types, crew members have special requirements. For example, the total crew number with day duty should be larger than a fixed value, which should be considered fully in the crew scheduling model. Furthermore, the crew number provided by bus company is limit, thus, the total crew number with all duty types must be no more than the crew number available.

The optimization objective of bus crew scheduling in this paper is to maximize the utilization of all crews, and consequently lower the operating cost. In order to make the crew more efficient, the objective function can be expressed by minimizing the total idle time of crew. The calculation of crew idle time can be shown in Figure 1. There are four trips assigned to a crew and the idle time section is three. The first idle time section begins with the arrival

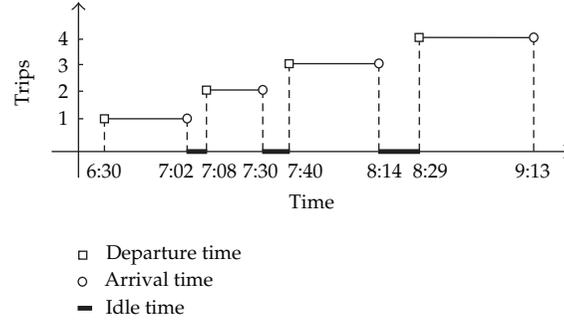


Figure 1: The crew idle time with four trips.

time of trip 1 (7:02) and ends with the departure time of trip 2 (7:08), and the value of first idle time is 6 minutes. Similarly, the second and third idle time is 10 and 15 minutes, respectively. Therefore, the total idle time of the crew is 31 minutes. The least idle time can be gained by arranging different trips with the same crew.

2.2. Definitions and Notations

The following definitions and notations are used throughout this paper.

N : number of trips provided by timetables in a day

G : number of duty types of crew

η : number of crew available in a day

θ : the least number of crew required with day duty

i : index of the trips, $i = 1, 2, \dots, N$

k : index of the crew, $k = 1, 2, \dots, \eta$

g : index of duty types, $g = 1$, represents early duty; $g = 2$, represents day duty; $g = 3$, represents late duty

z_g : number of crew with duty g

d_i : departure time of trip i at the starting station

a_i : arrival time of trip i at the terminal

T_0 : the least layover time at the starting station for each crew

T_g : the maximum value of work hours for each crew with duty g

B_g : the earliest sign-on time of crew with duty g

E_g : the latest sign-off time of crew with duty g

$x_{g,k}^i$: decision variables to indicate whether trip i is carried out by crew k with duty g , $x_{g,k}^i$ takes two values: $x_{g,k}^i = 1$, if trip i is carried out by crew k with duty g ; $x_{g,k}^i = 0$, otherwise

$y_{g,k}^{ij}$: a binary 0-1 variable which indicates the status of trip i and trip j carried out by crew k with duty g . $y_{g,k}^{ij}$ takes two values: $y_{g,k}^{ij} = 1$, if trip j is the next trip after trip i carried out by crew k with duty g ; $y_{g,k}^{ij} = 0$, otherwise.

2.3. Bus Crew Scheduling Model

As defined above, the crew scheduling problem can be formulated as follows:

$$\min \sum_{g=1}^G \sum_{k=1}^{z_g} \sum_{i=1}^N \sum_{j=1}^N y_{g,k}^{ij} \cdot (d_j - a_i), \quad (2.1)$$

$$x_{g,k}^j \cdot d_j - x_{g,k}^i \cdot a_i \geq T_0, \quad \forall g, k, \quad (2.2)$$

$$\sum_{i=1}^N x_{g,k}^i \cdot (a_i - d_i) \leq T_g, \quad \forall g, k, \quad (2.3)$$

$$B_g \leq x_{g,k}^i \cdot d_i \leq E_g, \quad \forall g, k, \quad (2.4)$$

$$z_2 \geq \theta, \quad (2.5)$$

$$\sum_{g=1}^G z_g \leq \eta, \quad (2.6)$$

$$\sum_{g=1}^G \sum_{k=1}^{z_g} x_{g,k}^i = 1, \quad \forall i, \quad (2.7)$$

$$y_{g,k}^{ij} = x_{g,k}^i \cdot x_{g,k}^j, \quad \forall g, k. \quad (2.8)$$

The objective is to minimize the total idle time of crew. As to crew k with duty g , the idle time between two adjacent trips i and j can be calculated by $y_{g,k}^{ij} \cdot (d_j - a_i)$. Inequality (2.2) guarantees that the time shift constraint should be satisfied by each of two adjacent trips carried out by the same crew. Inequalities (2.3) and (2.4) represent the constraints of work intensity and duty type compatibility, respectively. Inequality (2.5) is the number of day duty constraint, which denotes that the number of crew with day duty should be more than or equal to the constant θ . Inequality (2.6) is the crew number constraint, which means that the total number of crew required should be no more than the value η . Constraint (2.7) assures that each trip will be carried out by exactly one crew. Constraint (2.8) shows the relationship between variable $y_{g,k}^{ij}$ and decision variable $x_{g,k}^i$.

3. Algorithm Design

Tabu search algorithm uses a neighborhood search procedure to iteratively move from one potential solution to an improved solution, until some stopping criterion has been satisfied. The tabu search is a metaheuristic local search algorithm that can be used for solving

Crew	Duty	Trips			
①	△ ₁	1	5	7	10
②	△ ₁	2	4	9	
③	△ ₂	3	6	14	
④	△ ₂	8	11	13	
⑤	△ ₃	12	15		

Figure 2: The expression of solution.

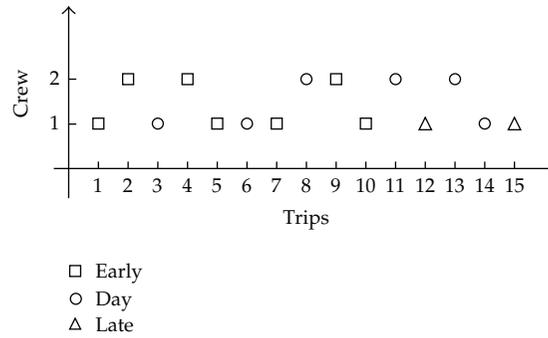


Figure 3: Decoding diagram.

combinatorial optimization problems. Because the major advantages of this algorithm are its simplicity, speed, and flexibility, and the urban bus crew scheduling model in the paper is a complex zero-one programming problem, we can use tabu search algorithm easily. The main parameters of algorithm are designed as follows.

3.1. Expression of Solution

The two-dimensional integer array encoding method can be used for the solution of crew scheduling problem, in which rows are the crews, the first column is the duty type of crew and other columns are the trips. The trips are numbered in ascending departure time order. Assume that there are fifteen trips carried out by five crews. The trip chains of each crew can be expressed as follows: crew 1: 1-1-5-7-10, crew 2: 1-2-4-9, crew 3: 2-3-6-14, crew 4: 2-8-11-13, and crew 5: 3-12-15; the expression of solution can be shown in Figure 2.

Based on the value of two-dimensional array, the decoding process is the inverse of encoding process. Take the data of Figure 2, as an example, there are three duty types of crew, and the crew number with early duty (duty 1) is two. The number of trips carried out by crew 1 with early duty is 1, 5, 7, and 10, thus, the variables are $x_{1,1}^1 = 1$, $x_{1,1}^5 = 1$, $x_{1,1}^7 = 1$, and $x_{1,1}^{10} = 1$. The number of trips carried out by crew 2 with early duty is 2, 4, and 9, thus, the variables are $x_{1,2}^2 = 1$, $x_{1,2}^4 = 1$, $x_{1,2}^9 = 1$. The decoding methods of other crews are similar to the former method. The decoding results are shown in Figure 3.

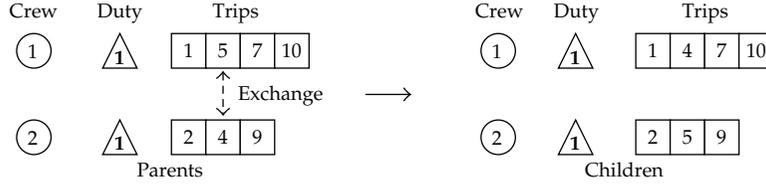


Figure 4: Trips exchange strategy.

3.2. Generation of Initial Solution

The initial solution is the starting point of algorithmic search. A superior quality of the initial solution will enable the algorithm to quickly converge to the optimal solution. In the process of generating the initial solution, the constraints with time shift, work intensity, and duty type compatibility should be satisfied. The procedure of the algorithm can be taken as follows:

Step 1 (Initialization). The trips set carried out by crew k with duty g is set to an empty set, namely, $\lambda_{g,k} = \emptyset$, for all g, k . The optional duty type set for trip i can be defined as follows: $Q_i = \{g \mid B_g \leq d_i \leq E_g, g \in \{1, 2, \dots, G\}\}$, for all i . Let $i = 1, z_g = 0$.

Step 2. Determine the duty type g of crew carrying out trip i with the following equation: $g = \min\{s \mid s \in Q_i\}$.

Step 3. Calculate the crew number k and the largest trip number i' corresponding to duty g . If $z_g > 0, k = \{l \mid \alpha_l = \min\{\alpha_l \mid l \in \{1, 2, \dots, z_g\}\}, \alpha_l = \max\{s \mid s \in \lambda_{g,l}\}\}, i' = \alpha_k$, go to Step 4; otherwise, go to Step 6.

Step 4. Verify time shift constraint. If $d_i - a_{i'} \geq T_0$, go to Step 5; otherwise, go to Step 6.

Step 5. Verify work intensity constraint. Calculate the total work time of crew k with duty g , $H_{g,k} = \sum_s x_{g,k}^s \cdot (a_s - d_s) + a_i - d_i$. If $H_{g,k} \leq T_g$, go to Step 7; otherwise, go to Step 6.

Step 6. Let $z_g \leftarrow z_g + 1, k \leftarrow z_g$, go to Step 7.

Step 7. Let $\lambda_{g,k} = \lambda_{g,k} \cup \{i\}, i \leftarrow i + 1$, go to Step 8.

Step 8. If $i > n$, stop and output the results; otherwise, go to Step 2.

3.3. Neighborhood Structure

The neighborhood structure uses trips exchange and inserts a strategy between different crews. Trips exchange strategy can be described as follows: a single exchange point on both two parents' trip chains is selected. The trip number of that point is swapped between the two parent organisms. The resulting organisms are the children. For example, trip 5 of crew 1 (1-1-5-7-10) is exchanged with trip 4 of crew 2 (1-2-4-9), and new solutions can be gained. The trips chain of crew 1 has become 1-1-4-7-10, and the trips chain of crew 2 has become 1-2-5-9, shown in Figure 4.

Trips inserting a strategy can be described as follows: one trip from crew 1 will be inserted into the trips chain of crew 2. The insert point depends on the ascending departure

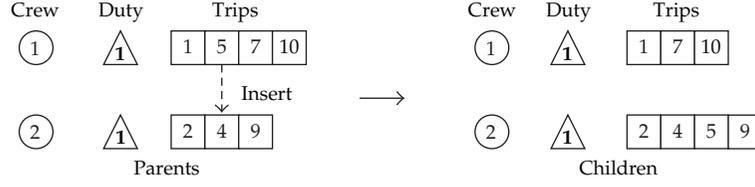


Figure 5: Trips insert strategy.

time order of the trips chain of crew 2, and two new trips chains of crew will reincarnation. For example, trip 5 of crew 1 (1-1-5-7-10) is inserted into the trips chain of crew 2 (1-2-4-9); new solutions can be gained. The trips chain of crew 1 has become 1-1-7-10, and the trips chain of crew 2 has become 1-2-4-5-9, shown in Figure 5. The trips exchange or insert operation cannot be carried out, if the trips chain of any crew cannot satisfy time shift, work intensity, or duty type compatibility constraints with any of the operations.

3.4. Evaluation of Solution

In order to search for better solutions in the algorithmic iterative process, it is necessary to evaluate the solution. It should calculate the value of objective function and consider the constraints at the same time. Because the initial solution has satisfied time shift, work intensity, and duty type compatibility constraints, and the new solution generated in neighborhood search also satisfies the former three constraints, the number of day duty and total crew constraints are the only factors which need to be considered. As for the solution which cannot satisfy the number of day duty and total crew constraints, it should be punished. Here we can take two parameters α and β as the punish factors, and both of the values take large positive numbers. If the solution can satisfy the number of day duty and total crew constraints, the value of fitness function and objective function will be equal; otherwise the value of fitness function will be much larger than the objective function, which means that the set of values for decision variables cannot create the feasible solution. The fitness function can be formulated as follows:

$$f = Z + \alpha \cdot \max\{\theta - z_2, 0\} + \beta \cdot \max\left\{\sum_{g=1}^G z_g - \eta, 0\right\}, \quad (3.1)$$

where $Z = \sum_{g=1}^G \sum_{k=1}^{z_g} \sum_{i=1}^N \sum_{j=1}^N y_{g,k}^{ij} \cdot (d_j - a_i)$ represents the value of objective function.

3.5. Other Parameters

The record of tabu table is the transform (exchange or insert) node, and tabu length takes fixed length. Select the regulation based on the value of evaluation as aspiration criterion, that is, the solution of the objective can be free, if it is better than any of the currently known best candidate solution. The stopping criterion is based on the value of fitness function. If the best value does not change after a given number of iterations, the algorithm will stop the calculation.

Table 1: Departure time of trips.

TN	DT	TN	DT	TN	DT	TN	DT	TN	DT	TN	DT
01	6:30	29	9:22	57	11:51	85	14:29	113	17:07	141	19:13
02	6:45	30	9:26	58	11:56	86	14:34	114	17:14	142	19:17
03	6:58	31	9:30	59	12:02	87	14:39	115	17:20	143	19:21
04	7:10	32	9:34	60	12:06	88	14:43	116	17:25	144	19:25
05	7:20	33	9:38	61	12:11	89	14:50	117	17:30	145	19:29
06	7:25	34	9:41	62	12:16	90	14:56	118	17:35	146	19:35
07	7:31	35	9:48	63	12:21	91	15:01	119	17:39	147	19:39
08	7:38	36	9:53	64	12:27	92	15:08	120	17:43	148	19:44
09	7:44	37	10:00	65	12:33	93	15:15	121	17:48	149	19:48
10	7:48	38	10:07	66	12:38	94	15:21	122	17:51	150	19:53
11	7:53	39	10:13	67	12:42	95	15:27	123	17:56	151	20:00
12	7:58	40	10:18	68	12:46	96	15:32	124	18:00	152	20:05
13	8:03	41	10:23	69	12:51	97	15:38	125	18:05	153	20:11
14	8:08	42	10:29	70	12:56	98	15:42	126	18:09	154	20:17
15	8:12	43	10:34	71	13:02	99	15:49	127	18:13	155	20:23
16	8:17	44	10:40	72	13:08	100	15:53	128	18:17	156	20:30
17	8:22	45	10:44	73	13:13	101	15:59	129	18:21	157	20:37
18	8:26	46	10:49	74	13:19	102	16:05	130	18:25	158	20:43
19	8:30	47	10:53	75	13:26	103	16:10	131	18:29	159	20:50
20	8:34	48	10:58	76	13:33	104	16:16	132	18:34	160	20:57
21	8:39	49	11:03	77	13:40	105	16:21	133	18:39	161	21:05
22	8:45	50	11:09	78	13:48	106	16:27	134	18:43	162	21:13
23	8:51	51	11:15	79	13:55	107	16:32	135	18:47	163	21:20
24	8:57	52	11:22	80	14:02	108	16:38	136	18:51	164	21:27
25	9:03	53	11:29	81	14:08	109	16:42	137	18:55	165	21:35
26	9:09	54	11:35	82	14:13	110	16:48	138	19:00	166	21:43
27	9:14	55	11:41	83	14:18	111	16:53	139	19:04	167	21:50
28	9:18	56	11:46	84	14:23	112	16:59	140	19:08	168	22:00

TN: trip number; DT: departure time.

4. Numerical Example

In a certain urban bus line, there are 168 trips in a day. The departure time of each trip can be obtained from Table 1. The running time from the starting station to the terminal is 30 minutes. The least layover time at starting station for each crew is 3 minutes. The maximum number of crew provided by bus company is 26, and the least crew number with day duty is 9.

The parameters of tabu search algorithm are taken as follows: the tabu length is 6, both the punish factors α and β are 10000, and the given number of iterations without improving the solution is 100. The other parameter is given in Table 2. The optimal solution can be calculated using VC++ program, shown in Table 3. The objective value of optimal solution is 1198 minutes.

Table 3 shows that the total number of crew is 26, where the crew number with early, day, and late duty is 8, 10, and 8, respectively. The earliest and latest departure times of each crew satisfy duty type compatibility constraint. Due to the same running time of trips, the crew with longest working time is the one who carries out with the largest number of trips,

Table 2: Parameter of different duty types.

Duty types	Earliest sign-on time	Latest sign-off time	Maximum working time (min)
Early	6:00	13:00	300
Day	9:00	18:00	300
Late	15:00	22:00	300

Table 3: The optimal results of crew scheduling.

Crew number	Duty type	Trips	Earliest departure time	Latest departure time
01	Early	1-4-9-16-24-31	6:30	9:30
02	Early	2-5-11-18-26-35-47-53-59	6:45	12:02
03	Early	3-7-14-22-30-39-46	6:58	10:49
04	Early	6-12-20-28-36	7:25	9:53
05	Early	8-15-23-32-40-48-54	7:38	11:35
06	Early	10-17-27-37-44	7:48	10:40
07	Early	13-21-29-38-45	8:03	10:44
08	Early	19-25-33-42-51-60-67	8:30	12:42
09	Day	34-41-52-61-68-74	9:41	13:19
10	Day	43-50-56-63-70	10:34	12:56
11	Day	49-55-62-69-75-80	11:03	14:02
12	Day	57-64-71-77-83-91-97-104-111	11:51	16:53
13	Day	58-65-72-78-84-90	11:56	14:56
14	Day	66-73-79-85-92-99-106	12:38	16:27
15	Day	76-81-88-95-102-109	13:33	16:42
16	Day	82-89-96-103-110	14:13	16:48
17	Day	86-93-100-108-114-121	14:34	17:48
18	Day	87-94-101-107-113	14:39	17:07
19	Late	98-105-112-118-127-135-143-151-157	15:42	20:37
20	Late	115-123-131-139-147-155-162-167	17:20	21:50
21	Late	116-124-132-140-148-156-164-168	17:25	22:00
22	Late	117-125-133-141-149	17:30	19:48
23	Late	119-129-137-145-153-161-166	17:39	21:43
24	Late	120-128-136-144-152-160-165	17:43	21:35
25	Late	122-130-138-146-154-159	17:51	20:50
26	Late	126-134-142-150-158-163	18:09	21:20

namely, the numbers 2, 12, and 19. The actual working time of the former three crews has the same value of 270 minutes, which is smaller than the maximum working time. The average crew working time with day duty is relatively small with the influence of time shift and the number of day duty constraints.

5. Conclusions

In this paper, a 0-1 integer programming model with the objective of minimizing the total idle time of crew for a circle bus line in consideration of mainly various working rules and duties

of different crews is developed. The considered working rules mainly involve time shift, work intensity, and duty type compatibility constraints, and the duties of crew include early, day, and late modes. A tabu search algorithm is then proposed to solve the model, and a numerical example has been given to show the effectiveness of the obtained results. The result shows that the model can effectively solve the scheduling problem with multiple duty types, and this method can thus serve as a useful tool for urban transit management department with more reasonable and pertinent assistant decision support. Furthermore, considering the crew scheduling problem with multiple transit lines and multiple depot constraints is an important topic for further research.

Acknowledgments

The work described in this paper was supported by National Natural Science Foundation of China (no. 50968009 and no. 71261014) and the Research Fund for the Doctoral Program of Higher Education (no. 20096204110003).

References

- [1] B. M. Smith and A. Wren, "A bus crew scheduling system using a set covering formulation," *Transportation Research Part A*, vol. 22, no. 2, pp. 97–108, 1988.
- [2] K. Darby-Dowman and G. Mitra, "An extension of set partitioning with application to scheduling problems," *European Journal of Operational Research*, vol. 21, no. 2, pp. 200–205, 1985.
- [3] M. Mesquita and A. Paias, "Set partitioning/covering-based approaches for the integrated vehicle and crew scheduling problem," *Computers and Operations Research*, vol. 35, no. 5, pp. 1562–1575, 2008.
- [4] J. E. Beasley and B. Cao, "A dynamic programming based algorithm for the crew scheduling problem," *Computers and Operations Research*, vol. 25, no. 7-8, pp. 567–582, 1998.
- [5] H. R. Lourenço, J. P. Paixão, and R. Portugal, "Multiobjective metaheuristics for the bus-driver scheduling problem," *Transportation Science*, vol. 35, no. 3, pp. 331–343, 2001.
- [6] D. Huisman, R. Freling, and A. P. M. Wagelmans, "Multiple-depot integrated vehicle and crew scheduling," *Transportation Science*, vol. 39, no. 4, pp. 491–502, 2005.
- [7] G. Mitra and K. Darby-Dowman, "A computer-based bus crew scheduling system using integer programming," *Computer Scheduling of Public Transport*, vol. 35, no. 5, pp. 223–232, 1985.
- [8] M. Desrochers and F. Soumis, "A column generation approach to the urban transit crew scheduling problem," *Transportation Science*, vol. 23, no. 1, pp. 1–13, 1989.
- [9] R. Clement and A. Wren, "Greedy genetic algorithms, optimizing mutations and bus driver scheduling," *Computer-Aided Transit Scheduling, Lecture Notes in Economics and Mathematical Systems*, vol. 430, pp. 213–235, 1995.
- [10] J. E. Beasley and B. Cao, "A tree search algorithm for the crew scheduling problem," *European Journal of Operational Research*, vol. 94, no. 3, pp. 517–526, 1996.
- [11] A. Ceder, *Public Transit Planning and Operation-Theory, Modelling and Practice*, chapter 10, Elsevier, Oxford, UK, 2007.
- [12] Y. Shen and Y. Ni, "Model for crew scheduling with time windows and multi-neighborhood structures," *Journal of Huazhong University of Science and Technology*, vol. 36, no. 12, pp. 31–34, 2008.
- [13] Y. J. Yang, Y. P. Wang, and X. M. Zhao, "Research on staff scheduling of urban passenger taxi dispatching center based on genetic algorithm," *Journal of Highway and Transportation Research and Development*, vol. 27, no. 7, pp. 142–146, 2010.
- [14] H. M. Niu, "Determination of the skip-stop scheduling for a congested transit line by bilevel genetic algorithm," *International Journal of Computational Intelligence Systems*, vol. 6, no. 4, pp. 1158–1167, 2011.
- [15] V. Srdjan, E. Aleksandar, and L. Imre, "Optimization of workflow scheduling in utility management system with hierarchical neural network," *International Journal of Computational Intelligence Systems*, vol. 4, no. 4, pp. 672–679, 2011.
- [16] O. Atli, "Tabu search and an exact algorithm for the solutions of resource-constrained project scheduling problems," *International Journal of Computational Intelligence Systems*, vol. 4, no. 2, pp. 255–267, 2011.

Research Article

Research on Public Transit Network Hierarchy Based on Residential Transit Trip Distance

Gao Jian, Zhao Peng, Zhuge Chengxiang, and Zhang Hui

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Zhao Peng, 11114242@bjtu.edu.cn

Received 9 May 2012; Revised 9 July 2012; Accepted 26 July 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Gao Jian 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 the problem of being lack of transit network hierarchy theory, a research on public transit network hierarchy optimization based on residential transit trip distance is conducted. Firstly, the hierarchy standard of transit network is given, in addition, both simulating electron cloud model and Rayleigh distribution model are used to fit the residential transit trip distance. Secondly, from the view of balance between supply and demand, the hierarchy step of transit network based on residential transit trip distance is proposed. Then, models of transit's supply turnover and demand turnover are developed. Finally, the method and models are applied into transit network optimization of Baoding, Hebei, China.

1. Introduction

Giving priority to the development of public transportation system plays a pivotal role in alleviating urban traffic congestion. Recently, with the fast expansion of the city size, the public transit lines get much larger and higher density. Many researchers attended to ensure each transit lines' function by hierarchical public transit network, and then a public transportation service system with clear hierarchy can be established, which can guide residence's travel behavior scientifically.

The notion of multilevel transit network planning has been widely acknowledged and adopted. Carrese and Gori [1] considered three-route hierarchies; the other studies focus on specific linkages such as feeder routes for a rail network; Bagloee and Ceder [2] divided the transit network into three degrees: mass route, feeder route and local route, and the method to determine hierarchy of a route is studied; Van Goeverden and van Nes [3] describes how the public transport system consists of different network levels; van Nes [4] proposed multilevel network optimization for public transport networks and he [5] also

conducted research on multiclass urban transit network design; Salzborn [6] and Knoppers and Muller [7] also optimized the transit network based on hierarchy concept. Jian and Gang [8], Wei et al. [9] proposed the multilevel transit network planning concept during planning the public transit network; Fangqiang [10] conducted transit network optimization based on transit network hierarchy from the view of coupling residential trip with transit network; Kuah and Perl [11] presented a mathematical model for feeder-bus network-design problem. They solved this model by a heuristic method called savings heuristic. Shrivastava and O'Mahony [12] designed feeders for one of heavy rail suburban service stations and coordinated schedules with the aid of genetic algorithm. Guillot [13] and Higgins [14] also conducted research on the bus network of a city was coordinated with the rail network. Chien et al. [15] applied genetic algorithm to design the feeders of a real network and its delay at intersections. Mohaymany and Gholami [16] used multiple modes with various capacities and performances in the feeder network design based on the minimization of user, operator, and social costs. Verma and Dhingra [17, 18] designed feeders of rail transit and presented a synchronized scheduling for rail and its feeders. Shrivastav and Dhingra [19] applied their heuristic feeder route generation algorithm to make a feeder network.

Although many researchers have conducted study related to hierarchy of public transit network, most of them did not propose any quantitative methods to calculate reasonable length for each hierarchical public transit network. In addition, travel behavior is widely studied and many transportation problems are solved based on it [20], so this paper attempts to put forward a hierarchy of public transit network which is applicable to certain urban development patterns and suitable for urban residents who travel by public transit in view of the fact that, at the moment, there is no such hierarchy based on the different demands of ridership and trip distance (resulting in the lack of arteries and local route lines).

2. Analysis on Transit Trip Distance Distribution

2.1. Method of Determining Hierarchy of the Transit Network

A rational public transit network should include different hierarchy types which operate with different standards. This paper categorizes public transit lines into 3 types: (1) mass route, which is the skeleton of the network; (2) feeder route, which operates inside of the district; (3) Local route, which serves as support facility to the mass route.

Hierarchy of a route and station spacing has been considered as the criteria of determining the network: (1) All the rail transit lines are regarded as mass routes, (2) public transit lines might qualify as mass routes when they are located along the express way or arterial road and the bus-stop spacing is longer than 800 m, (3) feeder route is the transit line that is located along the arterial road or subarterial road and the bus-stop spacing is between 500 m and 800 m, and (4) the remaining unclassified lines are all counted as local routes. There are still certain points of disagreements on the above criteria. Herein, we set up following provisions to pave the ways for later modeling.

- (1) One public transit line alone may cover different hierarchy types of route, as shown in Figure 1. In such case, if the trip covers different types, the transfer time is 0.
- (2) Lines passing the same section of the road are well considered to make sure that they all belong to the same hierarchy type.

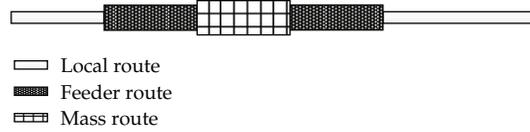


Figure 1: Sketch of transit network hierarchy types.

Table 1: Parameters and coefficients of determination in simulating electron cloud model and Rayleigh distribution model.

Cities	Electron cloud model		Rayleigh distribution model	
	a_0	R^2	λ	R^2
Shenyang	4.9	0.969	0.0312	0.962
Suzhou	7.7	0.977	0.0130	0.982
Qinhuangdao	6.2	0.974	0.0196	0.979
Bengbu	5.7	0.965	0.0233	0.965
Yinchuan	6.9	0.960	0.0160	0.953
Wujiang	5.1	0.977	0.0283	0.970
Changshu	4.6	0.963	0.0343	0.958
Suzhou	4.7	0.982	0.0341	0.979
Huaibei	4.5	0.980	0.0366	0.978
Changde	5.4	0.980	0.0258	0.977
Chaozhou	5.2	0.990	0.0275	0.985
Weifang	5.2	0.989	0.0270	0.983
R^2	Mean value = 0.976 Standard deviation = 0.009		Mean value = 0.973 Standard deviation = 0.010	

2.2. Analysis on Transit Trip Distance Distribution

The simulation of electron cloud model and Rayleigh distribution model are both widely used in the research of trip distance [16–18]. The applicability of these two models in the study of transit trip distance has been explored in this paper. Please refer to [17, 18] for more detailed introduction of these two models. Data from twelve different cities (Shenyang, Suzhou, Qinhuangdao, Bengbu, Yinchuan, Wujiang, Changshu, Suzhou, Huaibei, Changde, Chaozhou, and Weifang) have been used to compare the above models.

Cumulative probability distribution function of the electron cloud model and Rayleigh distribution model are: $F(s) = 1 - (2(s/a_0)^2 + 2(s/a_0) + 1)e^{-2(s/a_0)}$ and $F(s)1 - e^{(-0.5\lambda s^2)}$ respectively. Table 1 shows the values of parameters involved. The comparative analysis has been conducted in terms of precision in simulating and ability in interpreting:

- (1) Precision in simulating: values of fitting function R^2 indicate that both models have very high fitting precision. However, after analyzing the mean values and standard deviations of R^2 , the electron cloud model turned out to be a better approach.
- (2) Ability in interpreting: λ derived from the Rayleigh distribution model has no actual meaning, whereas a_0 derived from electron cloud model means the average trip distance. Consequently, the latter performs better.

Given those two aspects, electron cloud model is more feasible for the theory stated here.

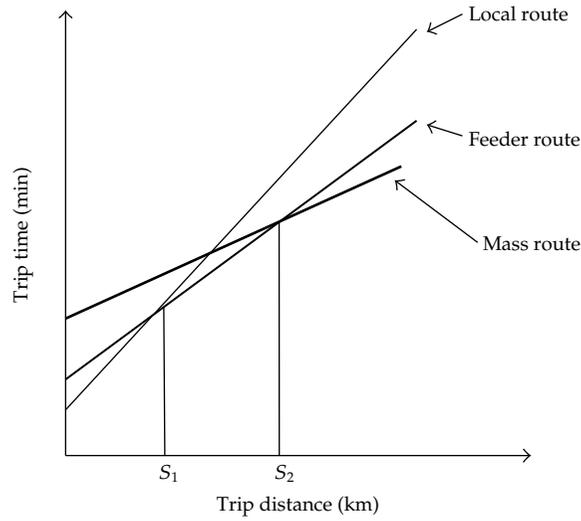


Figure 2: The relationship between transit trip distance and trip time.

3. Hierarchy of Transit Network Based on Balance between Supply and Demand

3.1. Concepts of Transit Hierarchy Planning

The optimized allocation of transit network is to study the balance between supply and demand of the public transit system at the macroscopic level. The typical quantitative indicators are the transit's supply turnover and demand turnover. Attempts have been made to balance the supply and demand through rational categorizing the network.

3.2. Analysis on Demand Turnover of Different Hierarchy Types

3.2.1. Optimal Trip Distance of Public Transit Lines in Different Types of Route

Different public transit lines have varies missions and desired length of passengers' trips (see Figure 2). When the trip distance is less than S_1 , local routes are more favorable. When it is between S_1 and S_2 , feeder routes are better options. If the distance is longer than S_2 , mass routes have obvious superiority. The problem of how to figure out preferential trip distance can thus be converted to the calculation of critical values S_1 and S_2 , which can easily be obtained by equations: $T_{\text{local}} = T_{\text{feeder}}$, $T_{\text{feeder}} = T_{\text{mass}}$ (T stands for the minimum trip time). Hence, the following model mainly focused on the trip time.

(I) Explanation and Hypothesis of the Model

- (1) Here, we assume that the layout of the system is rational, and typical square grid with proportional spacing was used, as shown in Figure 3.
- (2) The transfer of passengers follows a strict order: local route to feeder to mass, or the other way around.

- (3) Layout of every type of route is arranged in order of hierarchy (from mass to local), one encompasses another, that is, $r_3 > r_2 > r_1$, r_3, r_2 , and r_1 are spaces between mass routes (including express way and arterial road), feeders, and local routes.
- (4) Average transfer time equals waiting time, exclusive of time spent on walking to the transfer station. $T_{\alpha\beta.\text{transfer}}$ represents the average transfer time. ($\alpha, \beta = 1, 2, 3$ which stands for local, feeder, and mass, resp.).

(II) Modeling Procedure

Definition of the parameters involved are as follow:

$i = 1, 2, 3$ —local, feeder, mass;

$T_i = T_{i.\text{on foot}} + T_{i.\text{by bus}} + T_{i.\text{waiting}}$ —trip time using i as the highest type of route;

$T_{i.\text{on foot}}$ —total walking time from starting point to the station and from the station to destination;

$V_{\text{on foot}}$ —walking velocity;

$T_{i.\text{by bus}}$ —time spent on the vehicle;

$T_{i.\text{waiting}}$ —transfer time at transfer station;

A —area of the city;

L_k —length of each type of route; $k = 1, 2, 3$ —local, feeder, mass;

$\rho_k = L_k / A$ —road network density;

r_k —space between routes of the same hierarchy type;

$1/r_k = \rho_k / 2, r_k = 2A / L_k$;

it is assumed that d_k equals $r_k / 2$;

$\eta_{i.n}$ —coefficient of determination on transfer;

n —number of transfers;

if there is no transfer during the trip, $\eta_{i.n} = 0$, otherwise, $\eta_{i.n} = T_{\alpha\beta.\text{transfer}}$;

\bar{n}_i —average transfer time;

V_i —velocity of the vehicle on type of route i ;

P_{trip} —trip distance.

(1) *Minimum Trip Time Using Local Routes as the Highest Type of Route.* Figures 4 and 5 demonstrate the shortest path by this means. Span of travelling on foot is $(0, 2d_1]$ (see Figure 3), to simplify the model, mean value of d_1 is used.

The minimum trip time is

$$T_1 = T_{1.\text{on foot}} + T_{1.\text{by bus}} + T_{1.\text{waiting}}$$

$$T_{1.\text{on foot}} = \frac{d_1}{V_{\text{on foot}}}$$

$$T_{1.\text{by bus}} = \frac{(P_{\text{trip}} - d_1)}{V_1},$$

$$T_{1.\text{waiting}} = \eta_{1.1} + \eta_{1.2} + \cdots + \eta_{1.n} = \bar{\eta}_1.$$

(3.1)

(2) *Minimum Trip Time Using Feeder as the Highest Type of Route.* Figures 6 and 7 show the shortest path by this means. We assume that in this case, local routes only play a supplementary role, accordingly, the mean value of d_1 is still used as the travelling on foot, because the span of travelling by local routes is $(0, 2d_2]$ (see Figure 3). To simplify the model, mean value of d_2 is used.

The minimum trip time is

$$\begin{aligned}
 T_2 &= T_{2_on\ foot} + T_{2_by\ bus} + T_{2_waiting} \\
 T_{2_on\ foot} &= \frac{d_1}{V_{on\ foot}}, \\
 T_{2_by\ bus} &= T_{1_out} + T_{2_on\ the\ way} + T_{1_back} = \frac{d_2}{V_1} + \frac{(P_{trip} - d_1 - d_2)}{V_2} \\
 T_{2_waiting} &= \eta_{2.1} + \eta_{2.2} + \cdots + \eta_{2.n} = \bar{\eta}_2,
 \end{aligned} \tag{3.2}$$

where T_{1_out} is the time spent from starting point to feeder via local route. $T_{2_on\ the\ way}$ is time spent on the feeder. T_{1_back} is time spent from feeder to destination via local route. $T_{1_out} = T_{1_back}$.

(3) *Minimum Trip Time Using Mass Routes as the Highest Type of Route.* Figures 8 and 9 show the shortest path by this means. Assume that mass route is the major route taken, feeder and local route act as the supplement, can be obtained in the same way, the mean value of d_1 and d_2 are still used as the travelling on foot and traveling by local route, respectively. The span of travelling by feeder route is $(0, 2d_3 - 2d_1]$ (see Figure 3). To simplify the model, mean value of $d_3 - d_1$ is used.

The minimum trip time is

$$\begin{aligned}
 T_3 &= T_{3_on\ foot} + T_{3_by\ bus} + T_{3_waiting} \\
 T_{3_on\ foot} &= \frac{d_1}{V_{on\ foot}}, \\
 T_{3_by\ bus} &= T_{1_out} + T_{2_out} + T_{3_on\ the\ way} + T_{2_back} + T_{1_back} \\
 &= \frac{d_2}{V_1} + \frac{(d_3 - d_1)}{V_2} + \frac{[P_{trip} - (d_2 + d_3)]}{V_3}, \\
 T_{3_waiting} &= \eta_{3.1} + \eta_{3.2} + \cdots + \eta_{3.n} = \bar{\eta}_3,
 \end{aligned} \tag{3.3}$$

where T_{1_out} is the time spent from starting point to feeder via local route, T_{2_out} is the time spent from local route to mass route via feeder. $T_{3_on\ the\ way}$ is the time spent on the mass route. T_{2_back} is the time spent from mass route to local route via feeder. T_{1_back} is time spent from feeder to destination via local route. $T_{1_out} = T_{1_back}$, $T_{2_out} = T_{2_back}$.

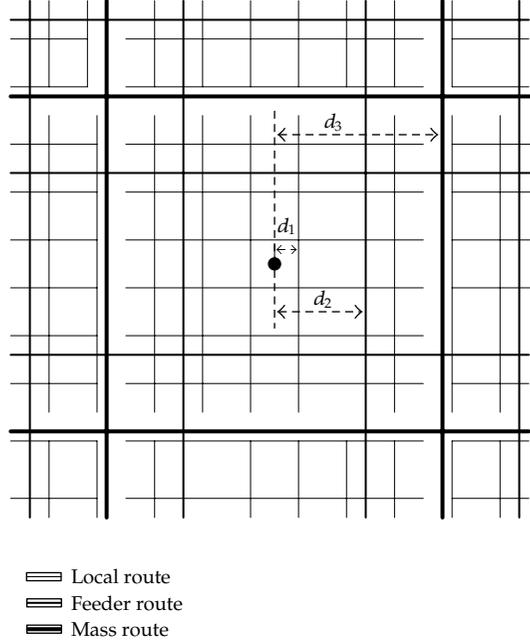


Figure 3: Sketch of transit network.

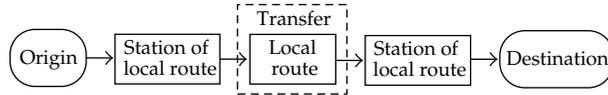


Figure 4: The shortest path using local routes as the highest type of route.

(4) Calculation of Optimal Distance on Every Type of Route. Set T_1 is equal to T_2 , T_2 , equal to T_3 , Then:

$$\begin{aligned}
 S_1 &= (d_1 + d_2) + \frac{V_2 \cdot V_1}{V_2 - V_1} \cdot (\bar{\eta}_2 - \bar{\eta}_1), \\
 S_2 &= (d_2 + d_3) + \frac{V_3 \cdot V_2}{V_3 - V_2} \cdot (\bar{\eta}_3 - \bar{\eta}_2).
 \end{aligned}
 \tag{3.4}$$

3.2.2. Analysis on Demand Turnover of Different Types of Route

(I) Proportion of Passengers on Each Highest Type of Route

The proportion of passengers who take local route, feeder, and mass route as their highest type of route is as follow.

Local Route: $w_1 = \int_0^{s_1} f(s)ds$; Feeder Route: $w_2 = \int_{s_1}^{s_2} f(s)ds$; Mass Route: $w_3 = \int_{s_2}^{+\infty} f(s)ds$.

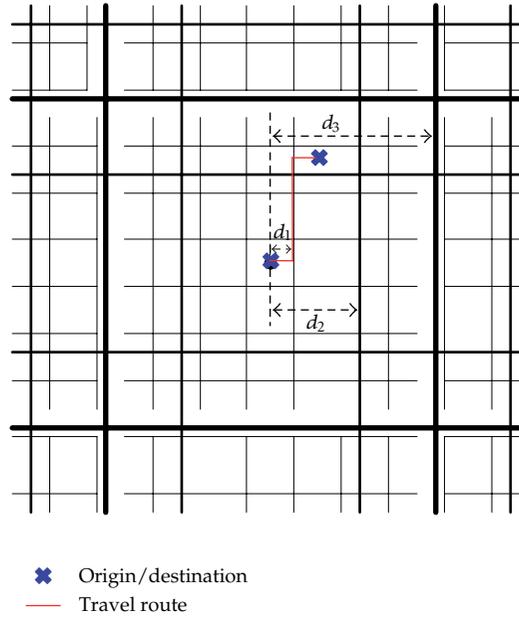


Figure 5: The sketch map of shortest path using local routes as the highest type of route.

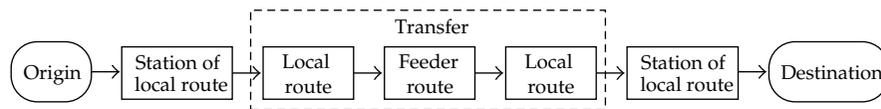


Figure 6: The shortest path using feeder as the highest type of route.

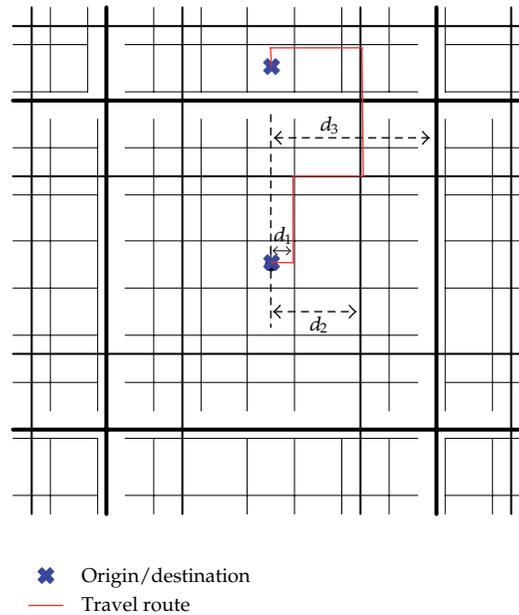


Figure 7: The sketch map of shortest path using feeder as the highest type of route.

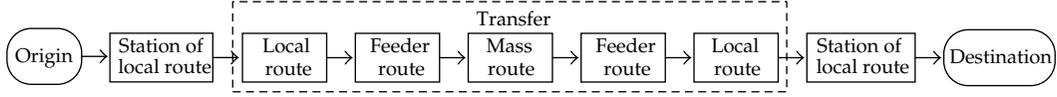
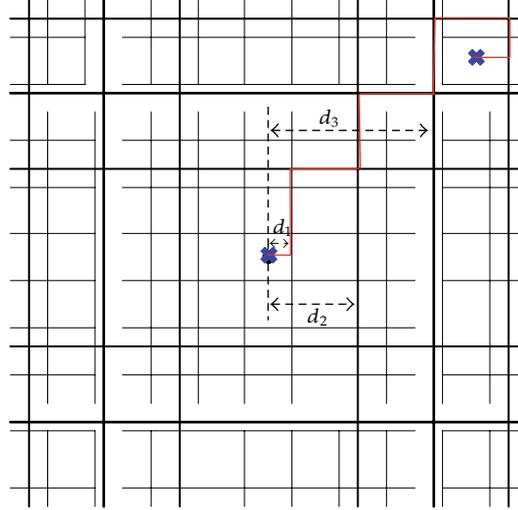


Figure 8: The shortest path using mass route as the highest type of route.



✕ Origin/destination
 — Travel route

Figure 9: The sketch map of shortest path using mass route as the highest type of route.

(II) Average Trip Distance of Passengers on Each Highest Type of Route

The average trip distance of passengers who take local route, feeder, and mass route as their highest type of route is as follow.

Local Route: $\bar{S}_1 = \int_0^{s_1} f(s)sd s / \int_0^{s_1} f(s)ds$; Feeder Route: $\bar{S}_2 = \int_{s_1}^{s_2} f(s)sd s / \int_{s_1}^{s_2} f(s)ds$;
 Mass Route: $\bar{S}_3 = \int_{s_2}^{+\infty} f(s)sd s / \int_{s_2}^{+\infty} f(s)ds$.

(III) Demand Turnover on Each Type of Route

Referring to previous studies [21, 22] on this issue, calculation method of demand turnover on each type of route is proposed as follows.

- (1) Z_i is defined as the turnover completed by unit passenger using i as the highest type of route, z_{ij} the component of Z_i completed on route type $j (j \leq i)$. Thus, $Z_i = \sum_{j=1}^i z_{ij}$.
- (2) z_{ij} is affected by the choice unit passenger make on each trip, which makes the solution complicated and time-consuming. Therefore, certain simplification has been made. We assume that the turnover on route type inferior to i can be seen as the product between the number of passengers who take i as the highest type of route and average trip distance made by passengers who take $i - 1$ as the highest

type. In this case, z_{ij} approximately is equal to Z_i minus the product. The rest such as turnover completed on route type inferior to $i-1$ can be done in the same manner.

Hence, demand turnover on each type of route can be calculated as follows.

- (1) Turnover completed on the local routes only

$$Z_1 = z_{11} = W \cdot w_1 \cdot \int_0^{s_1} f(s) ds, \quad (3.5)$$

where, W is the total number of trip time.

- (2) Turnover completed on the feeder routes only

$$\begin{aligned} Z_2 &= W \cdot w_2 \cdot \int_{s_1}^{s_2} f(s) ds = z_{21} + z_{22} \\ z_{21} &= \bar{S}_1 \cdot W \cdot w_2, \\ z_{22} &= Z_2 - z_{21} = W \cdot w_2 \cdot \left(\int_{s_1}^{s_2} f(s) ds - \bar{S}_1 \right). \end{aligned} \quad (3.6)$$

- (3) Turnover completed on the mass routes only

$$\begin{aligned} Z_3 &= W \cdot w_3 \cdot \int_{s_2}^{+\infty} f(s) ds = z_{31} + z_{32} + z_{33} \\ z_{31} &= \bar{S}_1 \cdot W \cdot w_3, \\ z_{31} + z_{32} &= \bar{S}_2 \cdot W \cdot w_3 \implies z_{32} = W \cdot w_3 \cdot (\bar{S}_2 - \bar{S}_1), \\ z_{33} &= Z_3 - (z_{31} + z_{32}) = W \cdot w_3 \cdot \left(\int_{s_2}^{+\infty} f(s) ds - \bar{S}_2 \right). \end{aligned} \quad (3.7)$$

- (4) Demand turnover on each type of route: demand turnover on local route, feeder; and mass route is represented as ZZ_1 , ZZ_2 , and ZZ_3 and total turnover Z_{total} . The relationships between Z_{total} , ZZ_i , Z_i , and z_{ij} are shown in Table 2.

3.3. Analysis on Supply Turnover of Different Hierarchy Types

Supply capacity of each type of route demands departure frequency, operating hours, type of vehicle, and load factor. Assume that supply turnover is GZ_i .

Departure Frequency: the average frequency on each type of route $f_i = 60/\bar{\mu}_i$ is used, where $\bar{\mu}_i$ is the average departure interval (min).

Table 2: The relationship among different classes (types) of turnover.

Hierarchy type	Local route	Feeder route	Mass route	Σ
Local route	Z_{11}	—	—	Z_1
Feeder route	Z_{21}	Z_{22}	—	Z_2
Mass route	Z_{31}	Z_{32}	Z_{33}	Z_3
Σ	ZZ_1	ZZ_2	ZZ_3	Z_{total}

Type of Vehicle: passenger flow varies on different types of route. Therefore, different types of vehicle are equipped accordingly. It is assumed that rated passenger load of vehicles on each type of route is E_i .

Load Factor: load factor ϕ_i is an important factor to indicate the comfort of the vehicle. Then, supply turnover of each type of route is

$$GZ_i = E_i \cdot T_{i,\text{run}} \cdot f_i \cdot \phi_i \cdot GL_i, \quad (3.8)$$

where $T_{i,\text{run}}$ is the operating hours in one day, GL_i is the length of each type or route.

3.4. Analysis on Balance between Demand and Supply on Each Type of Route

Based on the analysis in Sections 3.2 and 3.3, here, we assume that supply turnover equal demand turnover, then, $GZ_i = ZZ_i$ (see Table 2), the desired length of each type of route is.

$$GL_i = \frac{1}{E_i \cdot \phi_i \cdot T_{i,\text{run}}} \cdot ZZ_i. \quad (3.9)$$

4. A Case Study: Baoding

Located in Heibei Province, Baoding has 100 km² lands for construction and the population had reached 1.06 million by 2009. The length of express way, mass route, feeder, and local route are 82.6 km, 116.3 km, 74.1 km, and 222.7 km, respectively. The total length is 495.7 km. The application's process and results are as follows.

(1) Values of Parameters Involved

Based on analysis of the public transit survey of Baoding, the values of parameters involved (the detail explanation of the parameters can be found in Section 3.3) are calculated and the results are as follows:

- (1) $E = [E_1, E_2, E_3] = [72, 98, 98]$;
- (2) $\bar{\mu} = [\bar{\mu}_1, \bar{\mu}_2, \bar{\mu}_3] = [10, 8, 6]$;
- (3) $\bar{\eta} = [\bar{\eta}_1, \bar{\eta}_2, \bar{\eta}_3] = [5, 9, 12]$;
- (4) $V = [V_1, V_2, V_3] = [15, 20, 25]$;

Table 3: Road length of each transit hierarchy in baoding.

Different situations	Types of route			Total length (km)
	local (km)	feeder (km)	mass (km)	
$\varepsilon = 0.9$	1493	1229	347	3069
$\varepsilon = 0.95$	1576	1297	366	3239
$\varepsilon = 1.05$	1741	1434	405	3580
$\varepsilon = 1.10$	1824	1502	424	3751
Actual supply	399	645	690	1734

$$(5) \varphi_1 = \varphi_2 = \varphi_3 = 0.9;$$

$$(6) T_{1,\text{run}} = T_{2,\text{run}} = T_{3,\text{run}} = 16h.$$

(2) Important Outcomes

Trip distance is simulated with electron cloud model; the probability density function is $f(s) = (4s^2/(5.4)^3)e^{-2s/5.4}$, with $R^2 = 0.995$. The model's precision in simulating is high enough to be applied.

There exists a dynamic balance between supply of the transit network and demand of passengers. Dynamic balance coefficient is assumed as $\varepsilon(\varepsilon_i = GZ_i/ZZ_i)$. If $\varepsilon \leq 0.9$, supply is inadequate, if $0.9 < \varepsilon \leq 0.95$ or $1.05 < \varepsilon \leq 1.10$, supply just matches demand, if $0.95 < \varepsilon \leq 1.05$, supply matches demand perfectly, if $\varepsilon > 1.10$, supply is sufficient.

Table 3 shows the demand (obtained by calculation) and supply (actual data) undervaries situations. Comparative analysis indicates that the public transit network in Baoding has the following problems.

- (1) Total length is relatively short. Currently, the actual supply length of Baoding's transit line is 1734 km, in order to reach the level $\varepsilon = 0.9$ (the supply length should be 3069 km), nearly 1300 km length of transit line should be added. The main reason why the total length is so short is because the lack of local routes whose main function is to expand service range of the transit network and to make walking distance as short as possible. The actual supply length of local route is only 399 km, which is far from the demand of level (supply length of local route should be 1493 km), therefore, local routes should be relatively longer and have higher densities.
- (2) The length of mass routes is relatively long. Currently, the actual supply length of mass routes is 690 km, which is much longer (nearly 340 km longer) than the requirement of level $\varepsilon = 0.9$ (the supply length can be 347 km). Large-scale distribution centers and functional areas are connected by mass routes which require high-speed transport. However, currently the size of Baoding city is at moderate level, and trip distance of residents is generally short, the length of existing mass routes seems a bit redundant.

To solve these problems, the idea of hierarchy planning is proposed as follows.

- (1) Construction of local routes should be strengthened in order to shorten the distance between bus station and origin or destination, consequently shorten the walking distance which facilitates bus travel.

- (2) Increase or decrease the grade of the transit routes to achieve rational route configuration. For example, alter mass route to feeder route or feeder to local route.
- (3) According to the above two ideas, a specific measure is put out as an example to meet the demand of level $\varepsilon = 0.9$: (a) decrease the some mass routes' grade, thus, there will 350 km mass route will be changed into feeder routes, the length of feeder routes will be about 1000 km; (b) add another 1000 km of local routes. By the above two ways, the hierarchy of Baoding's public transit will be much reasonable and level $\varepsilon = 0.9$ can be reached.

5. Conclusion

A hierarchy planning toward public transit network is developed based on the distribution of passengers' trip distance. Main achievements are concluded as follows. (1) Trip distance is simulated with electron cloud model and Rayleigh distribution model, comparative analysis shows that the former has better precision in simulating and ability in interpreting. (2) A model for optimal trip distance of each hierarchy type of routes is proposed based on features of passengers in the public transit system; (3) A method of macroscopic calculation on hierarchy planning is developed, which is based on turnover balance between supply and demand. The above achievements have enriched the theory of hierarchy configuration of public transit network, and provide a feasible approach to transit network planning.

Acknowledgment

The authors would like to thank for financial support by the national science and technology support projects, under the Contract no. 2009BAG12A10-9.

References

- [1] S. Carrese and S. Gori, "An Urban bus network design procedure," *Applied Optimization*, vol. 64, pp. 177–196, 2002.
- [2] S. A. Bagloee and A. Ceder, "Transit-network design methodology for actual-size road networks," *Transportation Research Part B*, vol. 45, pp. 1787–1804, 2011.
- [3] C. D. van Goeverden and R. van Nes, "Hierarchy in public transport networks: the case of Amsterdam," in *Proceedings of the 11th World Conference on Transport Research*, 2007.
- [4] R. Van Nes, "Multilevel network optimization for public transport networks," *Transportation Research Record*, vol. 1799, pp. 50–57, 2002.
- [5] R. Van Nes, "Multiuser-class urban transit network design," *Transportation Research Record*, vol. 1835, pp. 25–33, 2003.
- [6] F. J. M. Salzborn, "Scheduling bus systems with interchanges," *Transportation Science*, vol. 14, no. 3, pp. 211–231, 1980.
- [7] P. Knoppers and T. Muller, "Optimized transfer opportunities in public transport," *Transportation Science*, vol. 29, no. 1, pp. 101–105, 1995.
- [8] L. jian and H. Gang, "Level planning method of bus-route network and its application," *Urban Transport of China*, vol. 2, no. 4, pp. 34–37, 2004.
- [9] W. Wei, Y. Xinmiao, and C. Xuewu, *Urban Transit Planning Method and Management Technology*, Science press, Beijing, China, 2002.
- [10] L. Fangqiang, *The Application of Coupling Models between Inhabitant Trip Distribution and Public Transport Network*, Southeast University, Nanjing, China, 2010.
- [11] G. K. Kuah and J. Perl, "The feeder-bus network-design problem," *Journal of the Operational Research Society*, vol. 40, no. 8, pp. 751–767, 1989.

- [12] P. Shrivastava and M. O'Mahony, "A model for development of optimized feeder routes and coordinated schedules-A genetic algorithms approach," *Transport Policy*, vol. 13, no. 5, pp. 413–425, 2006.
- [13] E. Guillot, "Bus transit interface with light rail transit in Western Canada," *Transportation Research Part A*, vol. 18, no. 3, pp. 231–241, 1984.
- [14] T. J. Higgins, "Coordinating buses and rapid rail in the San Francisco Bay Area: the case of Bay Area rapid transit," *Transportation*, vol. 10, no. 4, pp. 357–371, 1981.
- [15] S. I. Chien, L. N. Spasovic, S. S. Elefsiniotis, and R. S. Chhonkar, "Evaluation of feeder bus systems with probabilistic time-varying demands and nonadditive time costs," *Transportation Research Record*, no. 1760, pp. 47–55, 2001.
- [16] A. S. Mohaymany and A. Gholami, "Multimodal feeder network design problem: ant colony optimization approach," *Journal of Transportation Engineering*, vol. 136, no. 4, pp. 323–331, 2010.
- [17] A. Verma and S. L. Dhingra, "Feeder bus routes generation within integrated mass transit planning framework," *Journal of Transportation Engineering*, vol. 131, no. 11, pp. 822–834, 2005.
- [18] A. Verma and S. L. Dhingra, "Developing integrated schedules for urban rail and feeder bus operation," *Journal of Urban Planning and Development*, vol. 132, no. 3, pp. 138–146, 2006.
- [19] 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.
- [20] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research Part C-Emerging Technologies*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [21] S. Fei, *Research on Grade Proportion and Layout Method of Urban Road*, Southeast University, Nanjing, China, 2006.
- [22] Z. Zhuping, *Road Network Gradation Optimization Model According To Traffic Demand*, Southeast University, Nanjing, China, 2009.

Research Article

Evolving Model for the Complex Traffic and Transportation Network Considering Self-Growth Situation

Wei Zhang and Di Xu

School of Management, Xiamen University, Xiamen 361005, China

Correspondence should be addressed to Di Xu, dxu@xmu.edu.cn

Received 4 May 2012; Revised 29 June 2012; Accepted 13 July 2012

Academic Editor: Wuhong Wang

Copyright © 2012 W. Zhang and D. Xu. 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.

It has been approved that the scale-free feature exists in various complex networks, such as the internet, the cell or the biological networks. In order to analyze the influence of the self-growth phenomenon during the growth on the structure of traffic and transportation network, we formulated an evolving model. Based on the evolving model, we prove in mathematics that, even that the self-growth situation happened, the traffic and transportation network owns the scale-free feature due to that the node degree follows a power-law distribution. A real traffic and transportation network, China domestic airline network is tested to consolidate our conclusions. We find that the airline network has a node degree distribution equivalent to the power-law of which the estimated scaling parameter is about 3.0. Moreover the standard error of the estimated scaling parameter changes according to the self-growth probability. Our findings could provide useful information for determining the optimal structure or status of the traffic and transportation network.

1. Introduction

The study of the coevolution of the dynamics and the topology of transportation networks is a promising research topic, especially for the potential implications in infrastructures planning. Triggered by two invaluable papers [1, 2], relevant problems of the complex network have attracted a great deal of attention in the latest decade. Commonly cited examples cover various types of networks such as the information network [3], the social network [4], the communication network [5] and the traffic and transportation network [6–8]. Within the framework of the reference of [1], the scale-free network follows the power-law degree distribution. Several researchers discussed the scale-free characteristic of the traffic and transportation network, including the urban street network [3], the logistics network [9],

the transit network [10], and the airline network [6]. Using the static model, it was reported that the load distribution over traffic and transportation follows a power law [11, 12]. However, the previous works have done lots of assumptions and simplifications on the structure of traffic and transportation networks, such that there should be only one link between two nodes on the network [1, 13]. As it is known to all, the situation not only does exist, but also can be seen commonly that two nodes are connected by more than one link on the traffic and transportation network. For the street network of a city, we could easily imagine that there are many links connecting between one pair of nodes. If we use the adjacent matrix [6] to demonstrate the connecting degree between the two cities, the cell number in the matrix would apparently larger than one. However, these situations were not permitted in the previous researches of the complex network theory for the traffic and transportation network [1, 4, 13]. Furthermore, the evolution of the traffic and transportation network always has the “*self-growth*” characteristic. For example, the growth of the airline network always undergoes two ways. One way is to connect the exiting airport to the new constructed airport (the new added node) with new airlines. The other one is to open new airlines between the existing airports without any new airport constructed. The latter way of the evolution of the air network expresses the “*self-growth*” feature. Generally, the relevant decision makers of the traffic and the transportation departments frequently face the problems of opening new lines to service the exiting stations (or the airport, depots and, etc.) or constructing a new station or intersection when the network is expanded. The self-growth characteristic of the traffic and transportation network has not been elaborately described or dealt with in the existing literatures about the scale-free network.

It is true that the dynamics in the network adapts to the topology, but, on the other side, topology of the network just grows independently of the dynamics. It would be quite important to adding an adaptive mechanism for the growth of the network, where, for instance, new links are constructed to support congested routes. The idea of having each pair of nodes connected by more than one link is not just equivalent to having just one link with a higher capacity (this has been studied in many publications). The reason is that add more than one link between each pair of nodes could enrich the chance of route choice between a given origin-destination (OD) pair. Moreover, adding the link to connect two nodes could increase the structure complexity of the traffic and transportation network, which means the degree distribution of network, should be different from that of the networks of the previous researches.

In this paper, a network evolving model for the traffic and transportation network is designed. In this evolving model, we consider the self-growth phenomenon during the evolving of the network. Based on the new network evolving mechanism, the node degree distribution of the generated network is deduced. From the power-law fitting results of the cumulative distribution functions (CDF) of the node degree distribution, we conclude that the traffic and transportation network owns the scale-free feature. However, the scale-free characteristic depends much on the parameters of self-growth during the evolution of the network.

2. Evolving Model for the Traffic and Transportation Network

Absenting from the classical random network models [1], two mechanisms are responsible for the emergence of the scaling in the exiting literatures: (1) the network evolves by adding a new node with m ($m \leq m_0$) edges which links the new node to m different nodes in

the primary small network (with m_0 numbers of nodes); (2) the preferential attachment which specifies the probability that a new node will be connected to node i depends on the degree of node i , that is, $\pi_i(k_i) = k_i / \sum_j k_j$. As has been stated in the first section, for the traffic and transportation network, they are commonly existed more than one link between two nodes. Furthermore, the evolution of traffic and transportation network always has the “self-growth” characteristic. Consequently, we design a new evolving model adaptive to the traffic and transportation network as follows.

Starting with a small number (m_0) of nodes, at each time step $t = 1, 2, \dots, T$, repeat the following operations $m(\leq m_0)$ times.

- (1) Randomly select a node in the current network as the *ending point* of the new link, and whether node i is selected as the ending point of the new link depends on its degree k_i with the preferential probability $\pi_i(k_i) = k_i / (\sum_j k_j)$.
- (2) Generate a random number $r(0 \leq r \leq 1)$, if $r \leq \rho$ (ρ denotes the *self-growth probability* for the evolution of the network), select randomly a node $j(j \neq i)$ in the system as the *starting point* of the new added link (j, i); otherwise if $r > \rho$: check if a new node k has been added into the system, if not, add a new node k to the system as the starting point and connect the ending point i to the new link (k, i), if yes, connect the new node k to the ending point i to add the new link (k, i).

Apparently, at each time step t, m new links can be added to the system regardless of that whether a new node is added. In order to demonstrate the growth mechanism for the traffic and transportation network, we use Figure 1 as illustration.

The primary network is shown in Figure 1(a), in which the number of links $m_0 = 4$ and the degree of each node in the system is respectively 1, 4, 2 and 1. Assume $m = 4$ and $\rho = 0.5$, at time $t = 1$, we should run $m(m = 4)$ times of the following operations to add four links into the system.

Step 1. Based on preferential probability $\pi_i = k_i / \sum_j k_j$, select randomly a node (assume node #2) as the ending point of new link.

Generate a random number r (assume $r = 0.2$). Since $r(= 0.2) \leq \rho(= 0.5)$, select randomly a node (assume node #3) in the system as the starting point to add the new link (3, 2).

Step 2. Like Step 1, select randomly a node (assume node #2) as the ending point of new link.

Generate a random number r (assume $r = 0.7$). Since $r(= 0.7) > \rho(= 0.5)$, and there has not been a new node added into the system, add a new node (node 5) to the system and connect it to the ending point (node 2) to add a new link (5, 2).

Step 3. Select randomly a node (assume node 4) as the ending point of new link. Generate a random number r (assume $r = 0.6$). Since $r(= 0.6) > \rho(= 0.5)$, and there has been a new node (number 5) in the system, connect node 5 and node 4 to add the new link (5, 4).

Step 4. Select randomly a node (assume node #2) as the ending point of new link. Generate a random number r (assume $r = 0.65$). Since $r(= 0.65) > \rho(= 0.5)$, and there has been a new node (number 5) in the system, connect node 5 and node 2 to add the new link (5, 2).

As illustrated in Figure 1(b), we are clear about that during the evolving of the traffic and transportation network, the network could grow by adding new links between the existing nodes or adding a new node and connect it to the existing nodes. Furthermore, each pair of nodes could be connected with more than one link. Given the time period $t = 1, \dots, T$,

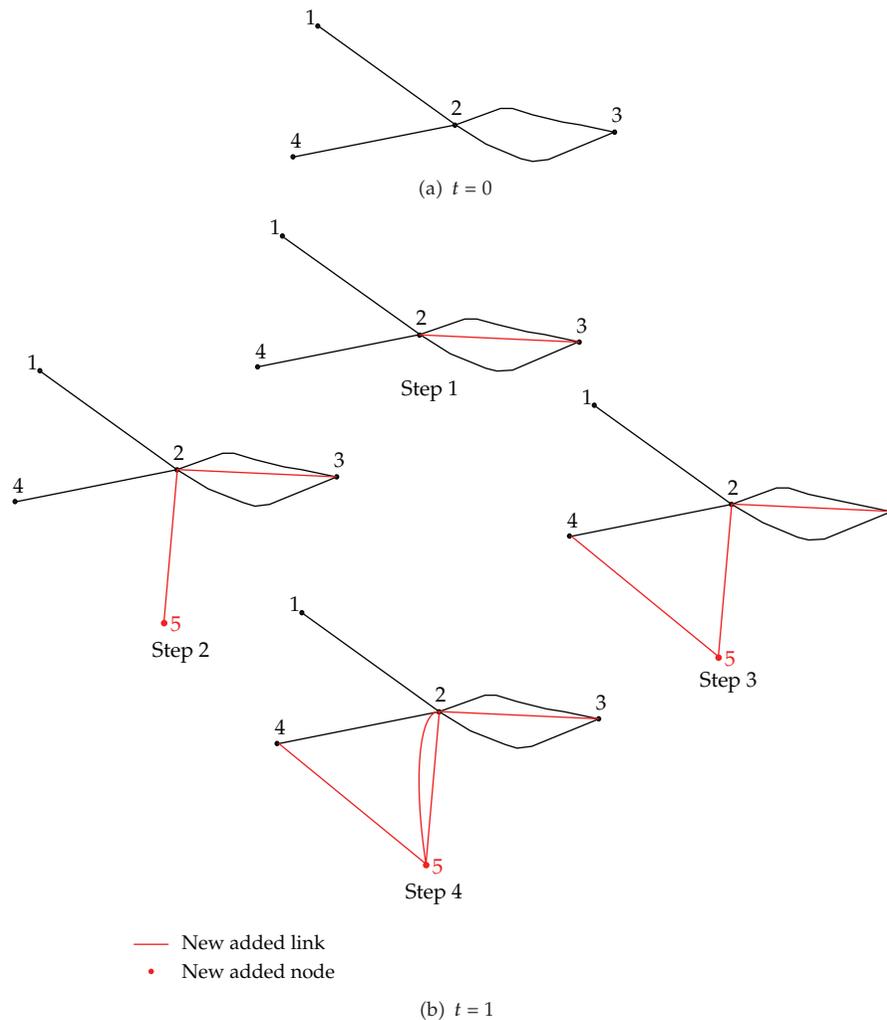


Figure 1: An illustration of the network evolving mechanism for the traffic and transportation network: Assume $m_0 = 4$, $m = 4$, and $\rho = 0.5$.

as well as the self-growth probability ρ , we could generate the traffic and transportation network which owns the self-growth feature.

3. Degree Distribution Model of the Traffic and Transportation Network

According to the continuum theory, we assume that the degree of node i changes continuously versus the time step t , denoted by $k_i(t)$. Let $N(t)$ represent the total number of nodes and let $L(t)$ be the total number of links in the system at time t . Start with $t = 0, N(t = 0) = m_0$ nodes, and $L(t = 0) = l_0$ links. $k_i(t)$ satisfies the following dynamical

equation:

$$\begin{aligned}
\frac{\partial k_i}{\partial t} &= (1 - \rho^m)m\pi_i(k_i) + \rho^m m \left[\pi_i(k_i) + (1 - \pi_i(k_i)) \frac{1}{N(t) - 1} \right] \\
&= m\pi_i(k_i) + \rho^m m \left[(1 - \pi_i(k_i)) \frac{1}{N(t) - 1} \right] \\
&= m \frac{k_i}{\sum_j k_j} + \rho^m m \left[\left(1 - \frac{k_i}{\sum_j k_j} \right) \frac{1}{N(t) - 1} \right].
\end{aligned} \tag{3.1}$$

In (3.1), the first term $mk_i/(\sum_j k_j)$ represents the probability that node i is selected as the ending point of the new added link at time step t in the standard BA model [1, 4]. As stated in the above section, at each time step of the evolving of traffic and transportation network, we would run m times of judging whether a new node would be added into the system. The probability that no node is added into the existing system is ρ^m (ρ denoted as self-growth probability). So, the event that at least one node is added into the system happens with a probability of $(1 - \rho^m)$, which multiplies $mk_i/(\sum_j k_j)$ to obtain the probability that node i is selected as the ending point and linked to a new added node at time t . Similarly, the second term $\rho^m m[(1 - k_i/(\sum_j k_j))(1/(N(t) - 1))]$ represents the probability that node i is selected as the starting point of the new added link in case that no node is added into the system (that is, the network undergoes a self-growth process).

According to [4], the sum of the degrees of all nodes is $\sum_j k_j = 2mt$. Accordingly, expected total number of the nodes in the system depends on the self-growth probability ρ , that is $N(t) = m_0 + (1 - \rho^m)mt$. Consequently, (3.1) could be written as

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} + \rho^m m \left[\left(1 - \frac{k_i}{2mt} \right) \frac{1}{m_0 + (1 - \rho^m)mt - 1} \right] \tag{3.2}$$

$$\approx \frac{k_i}{2t} + \rho^m m \left[\left(1 - \frac{k_i}{2mt} \right) \frac{1}{(1 - \rho^m)mt} \right], \quad \text{for large } t,$$

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} + \frac{\rho^m(2mt - k_i)}{2(1 - \rho^m)mt^2} \tag{3.3}$$

$$= \frac{k_i}{2t} + \frac{\rho^m}{(1 - \rho^m)t} + \frac{\rho^m k_i}{2(1 - \rho^m)mt^2} \tag{3.4}$$

$$\frac{\partial k_i}{\partial t} \approx \frac{k_i}{2t} + \frac{\rho^m}{(1 - \rho^m)t}, \quad \text{for large } t.$$

The solution of (3.4), with the initial condition that node i was added to the system at time t_i with the degree $k_i(t_i) = m(1 - \rho^m)$ (on the condition that no self-growth event has happened), is

$$k_i(t) = A \left(\frac{t}{t_i} \right)^{1/2} - \frac{2\rho^m}{(1 - \rho^m)}, \quad \text{with } A = \frac{(m - 2m\rho^m + 2\rho^m + m\rho^{2m})}{1 - \rho^m}. \tag{3.5}$$

Using (3.5), the probability that a node at equal time has a degree k_i smaller than k , $P(k_i(t) < k)$, could be expressed as:

$$P(k_i(t) < k) = P(t_i > c(k)t) \quad (3.6)$$

Based on (3.5) and (3.6), we have

$$c(k) = \left(\frac{A(1 - \rho^m)}{k(1 - \rho^m) + 2\rho^m} \right)^2. \quad (3.7)$$

Simply, t_i follows a constant probability density as:

$$P(t_i) = \frac{1}{m_0 + t_i} \quad (3.8)$$

Substituting this into (3.6), we could obtain

$$P(t_i > c(k)t) = 1 - \frac{c(k)t}{m_0 + t} = 1 - \frac{t}{m_0 + t} \left(\frac{A(1 - \rho^m)}{k(1 - \rho^m) + 2\rho^m} \right)^2. \quad (3.9)$$

$P(k)$ could be obtained using:

$$\frac{\partial P(k_i(t) < k)}{\partial k} = \frac{\partial P(t_i > c(k)t)}{\partial k} = \frac{2t}{m_0 + t} \left(\frac{A^2(1 - \rho^m)^3}{(k(1 - \rho^m) + 2\rho^m)^3} \right) \quad (3.10)$$

So that, we have:

$$\begin{aligned} P(k) &= \frac{\partial P(k_i(t) < k)}{\partial k} \sim \left[\frac{(m - 2m\rho^m + 2\rho^m + m\rho^{2m})}{1 - \rho^m} \right]^2 \frac{2(1 - \rho^m)^3}{[(1 - \rho^m)k + 2\rho^m]^3}, \quad t \rightarrow \infty \\ &= \frac{2(1 - \rho^m)(m - 2m\rho^m + 2\rho^m + m\rho^{2m})^2}{[(1 - \rho^m)k + 2\rho^m]^3}. \end{aligned} \quad (3.11)$$

From (3.11), we know that the degree distribution of the traffic and transportation network depends on the self-growth probability ρ . Clearly, (3.11) demonstrates that $P(k)$ follows a power-law distribution with scaling parameter (exponent gamma) near 3. Consider the following two cases:

- (1) in case that $\rho \rightarrow 0$, we have $P(k) \rightarrow 2m^2k^{-3}$. The evolving model is equivalent to the standard BA model [13], which has the scale-free feature according to [1, 2].
- (2) in case that $\rho \rightarrow 1$, $P(k) \rightarrow 0$, which implies that the network is not growing in the way of adding new nodes at all. However, even in this case using the static model in [12], one can still yield scale-free $P(k)$.

4. Simulation Results

To provide numerical support for the degree distribution function of (3.11), we generate different networks of $N = 10^4$ nodes with initial configurations shown in Figure 1(a). We set $m=4$ versus different self-growth probabilities $\rho=0.0, 0.1, 0.3, 0.5, 0.7,$ and 0.9 , respectively. After performing the network evolving model in Section 2 (set $T = N = 10^4$), the node degree distribution result of each network is shown in Figure 2.

In Figure 2, we plot the histograms of event frequency versus size, event rank versus size, and event rank versus frequency of the node degree of the generated networks in panel (a) to (c) to panel (c). We also examine the log-normal behavior of the node degree distribution using the method in [11, 14], and. The result is shown in panel (d).

We could see that these curves on panel (a) to panel (c) display the situations of “long tails,” which is the reprehensive nature of power-law distributing data set. On the other hand, as shown in panel (d) of Figure 2, it could observed that the node degree data sets own lognormal distribution behaviors in the fact that distance from means of log degree values versus square roots of logs of counts of binned values have leaner relationships according to the regression (the method used for the leaner regression could be referred in [4, 11, 14]). Accordingly, we could say again that our data set of node degree does have power-law distribution nature.

5. Applications to Real-World Data

To consolidate our result, we investigate the evolving characteristics of real traffic and transportation networks. We investigate the evolution of China domestic airline network (CDAN) from the year 1950 to 2010 and 2000 to 2010 for detailed network topology information. The node degree datasets of the networks were collected in May, 2010 from the database of the Civil Aviation Bureau of China [15].

We summarized average node degrees, number of airlines and airports (nodes) of these networks of CDAN from 1950 to 2010, which are shown in Figure 3. As described in Figure 3, we found that the average degree of networks of CDAN increased drastically from year 1950 to 2000. However, the increase of the average node degree is relatively stationary from 2000 to 2010. The change trend of the number of airports in the network was the same as that of the average node degree. However, there was somewhat difference in change trend of the number airline of CDAN.

As shown from the green diamond line in Figure 3, we can see that the increase rate of the number of airlines is more sharply than those of the rest two items from the year 2000 to 2010. With this evidence, we could conclude that the evolution of CDAN from 2000 to 2010 obviously undergo the process of self-growth. Indeed, the number of connections between any two cities of China with the airline has increased more than 50 lines per year since 2000. On the other hand, there were only 42 new airports opened in the recent decade [15].

For a better and more complete picture of the self-growth evolution of CDAN, we plot the cumulative distribution function (CDF) of the node degree of the network of 2000, 2005, and 2010 in Figure 4. In Figure 4, we also display straight line of the fitting of power-laws of the CDF of each the node degree datasets using the MLE using in [11]. Apparently, in the three panels, the node degree data set fit the logarithmic power-law line so close that we also could know that CDAN owns the scale-free feature. From Figure 4, it could be clearly recognized that the data sequence of node degree of the network does not arrange on

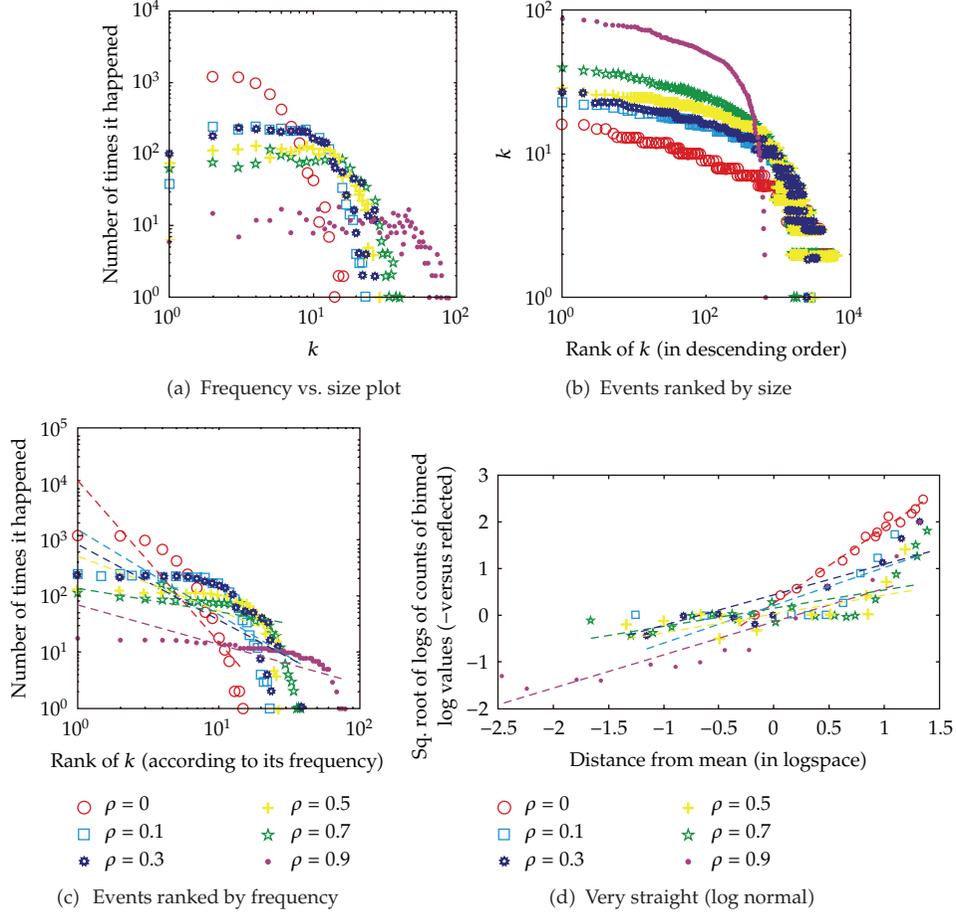


Figure 2: Histograms of event frequency versus size, event rank versus size, event rank versus frequency and the log-normal behavior test of the node degree of networks generated with the proposed evolving model with $m = 4$ versus different self-probabilities.

a straight line as the discrete power-law distributing data (which is as shown in [11]). This situation tells that the previous reports (which reported the node degree follows the pure power-law distribution, e.g., [16]) could not perfectly unfold the scaling-free feature of the airline network. As evident in Figure 4, it could be seen that, for CDAN the number of nodes with a degree more than 10 is larger and larger from year 2000 to 2010. This situation implies that the network has evolved mainly in the way of connecting the existing airports with the newly opened airlines from year 2000 to 2010, rather than opening the new airport.

Since the data node degrees of CDAN from 2000 to 2010 has been approved following the power-law distribution. For measuring the qualities of these distributions, we test the power-law hypothesis using the methods described in [11]. The relevant results are summarized in Table 1, in which: $\hat{\alpha}$ denotes the estimated value of the scaling parameters of power-law distribution; $\sigma_{\hat{\alpha}}$ is the standard error on $\hat{\alpha}$; $\langle k \rangle$ is the average node degree as mentioned before; k_{\min} is the smallest node degree.

As shown in Table 1, we could see in the second column that the estimated scaling parameters $\hat{\alpha}$ of power-law distribution of the node degree of CDAN decreases gradually

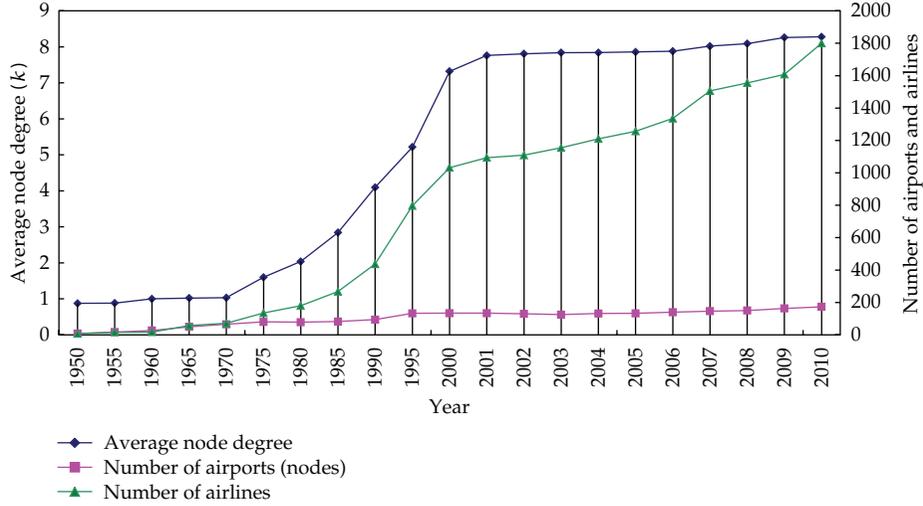


Figure 3: Average node degree and the number of airports/airlines of the China domestic airline network from year 1950 to 2010.

Table 1: Basic parameters of power-law distribution tests of China domestic airline networks, from year 2000 to 2010.

Year	$\langle k \rangle$	$\hat{\alpha}$	$\sigma_{\hat{\alpha}}$	k_{\min}
2000	7.26	3.09	3.12	1.00
2001	7.32	3.15	4.45	1.00
2002	7.76	3.21	6.51	1.00
2003	7.81	2.45	6.75	1.00
2004	7.84	2.35	7.18	1.00
2005	7.84	2.46	8.52	1.00
2006	7.86	2.33	8.19	1.00
2007	7.88	2.22	9.62	1.00
2008	8.02	2.11	11.25	1.00
2009	8.09	2.15	17.01	1.00
2010	8.26	2.12	19.15	1.00

from 2000 to 2010. Instead, the standard error on $\hat{\alpha}$ increases bit by bit from 2000 to 2010. This phenomenon could be explained by the self-growth nature of the airline network. From year 2000 to 2010, the self-growth probability of the network might be relevantly small initially. Consequently, the estimated scaling parameter of the power-law distribution is closed to 3.0 with a small standard error. However, as the self-growth probability increases, the estimated scaling parameter of power-law distribution as well as its standard error is heavily influenced by the self-growth situation. This situation could also be interpreted using (3.11). If the self-growth probability ρ is large enough and closed to 1.0, the node degree distribution of the network might not follow the pure power-law distribution. That is the estimated scaling parameter of power-law distribution would be smaller than 3.0 and have the larger standard error.

Comparing with the published works, it was reported in [10, 17, 18] that traffic and transportation networks followed power-law distributions and hold a scaling parameter

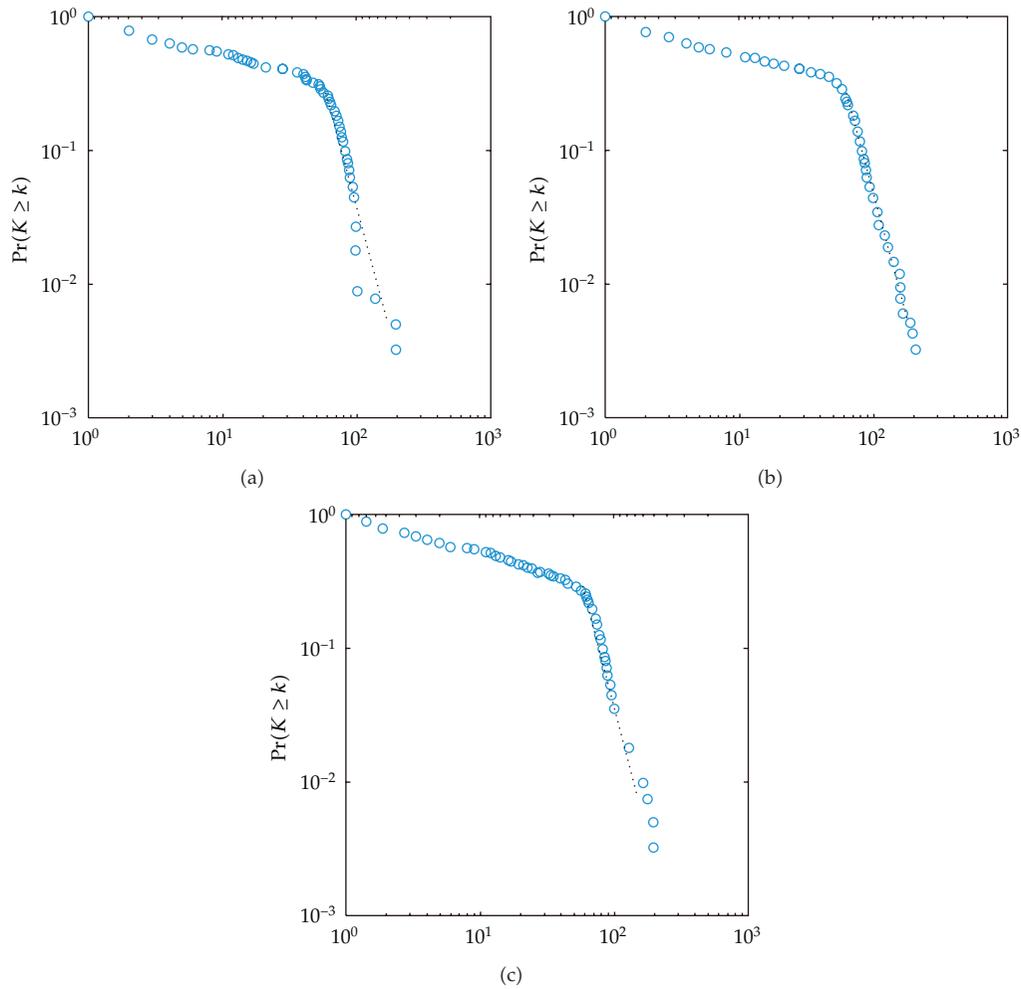


Figure 4: The cumulative distribution function $\Pr(K)$ and their straight line power-law fits of the node degree of China Domestic airline network at year 2000, 2005, 2010.

range from 1.0 to 2.4. Here, we approve that CDAN is scale-free but follows a power-law distribution with a scaling parameter between 3.0 with the varied standard error, which is constrained by the self-growth parameters during the evolution of the airline network.

6. Conclusions

The study of the structure of the traffic the transportation network from the perspective of complex network has drawn considerable researching attention. Unfortunately, the opinions of the previous researchers are identical that they take account of the evolution of the traffic and transportation network without considering the situation of self-growth. In this paper we proposed an evolving model which is relevantly adaptive for the traffic and transportation network. According to the analytic results for node degree distributions of the tested networks generated via the proposed model, we find that the networks are scale-free

due to that their node degree distributions follow the power-law distributions. We also test the distribution of the node degree of the real traffic and transportation network and conclude that our evolving model is more appropriate to explain the scale-free feature as well as the self-growth situation of the traffic and transportation network.

We hope that the methods given in this paper could provide useful evidences for determining the optimal structure of the traffic and transportation network, which also is crucial to our future research. Geographic or spatial networks, such as an urban street network or railway network, have strong constraints during their growth due to the fact that their nodes and edges are embedded in time and space. This fact should also be explicitly investigated in the future research.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant no. 71171171. The authors would like to thank anonymous reviewers for their insightful comments and suggestions.

References

- [1] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [2] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [3] A. L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nature Reviews Genetics*, vol. 5, no. 2, pp. 101–113, 2004.
- [4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [5] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 11, pp. 3747–3752, 2004.
- [6] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, "The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 22, pp. 7794–7799, 2005.
- [7] 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.
- [8] J. J. Wu, Z. Y. Gao, and H. J. Sun, "Model for dynamic traffic congestion in scale-free networks," *Europhysics Letters*, vol. 76, no. 5, p. 787, 2006.
- [9] B. Tadić, S. Thurner, and G. J. Rodgers, "Traffic on complex networks: towards understanding global statistical properties from microscopic density fluctuations," *Physical Review E*, vol. 69, no. 3, Article ID 036102, 5 pages, 2004.
- [10] H. Sun, J. Wu, and Z. Gao, "Dynamics of traffic networks: from microscopic and macroscopic perspectives," *Physica A*, vol. 387, no. 7, pp. 1648–1654, 2008.
- [11] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.
- [12] K. I. Goh, B. Kahng, and D. Kim, "Universal behavior of load distribution in scale-free networks," *Physical Review Letters*, vol. 87, no. 27, Article ID 278701, 4 pages, 2001.
- [13] M. Barthélemy, "Spatial networks," *Physics Reports*, vol. 499, no. 1–3, pp. 1–101, 2011.
- [14] S. Milojević, "Power law distributions in information science: making the case for logarithmic binning," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2417–2425, 2010.
- [15] China and C. A. A. O., *Statistical Data on Civil Aviation of China 1950–2010*, Chinese Civil Aviation Press, 2010.

- [16] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 21, pp. 11149–11152, 2000.
- [17] R. V. Donner, J. Heitzig, J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "Geometric detection of coupling directions by means of inter-system recurrence networks," *Physics Letters A*, vol. 373, no. 46, pp. 4246–4254, 2009.
- [18] H. Jeong, "Complex scale-free networks," *Physica A*, vol. 321, no. 1-2, pp. 226–237, 2003.

Research Article

Handling Stability of Tractor Semitrailer Based on Handling Diagram

Ren Yuan-yuan,^{1,2} Zheng Xue-lian,² and Li Xian-sheng²

¹ College of Computer Science and Technology, Jilin University, No. 2699 Qianjin Street, Changchun 130012, China

² Traffic College, Jilin University, No. 5988 Renmin Street, Changchun 130022, China

Correspondence should be addressed to Li Xian-sheng, lixiansheng123@yahoo.com.cn

Received 9 May 2012; Revised 9 July 2012; Accepted 18 July 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Ren Yuan-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.

Handling instability is a serious threat to driving safety. In order to analyze the handling stability of a tractor semitrailer, a handling diagram can be used. In our research, considering the impact of multiple nonsteering rear axles and nonlinear characteristics of tires on vehicle handling stability, the handling equations are developed for description of stability of tractor semi-trailer. Then we obtain handling diagrams so as to study the influence of driving speed, loaded mass, and fifth wheel lead on vehicle handling stability. The analysis results show that the handling stability of a tractor semi-trailer when the tractor has two nonsteering rear axles is better than that when the tractor has only one nonsteering rear axle. While the stability in the former case is slightly influenced by driving speed and loaded mass, the latter is strongly influenced by both. The fifth wheel lead is found to only slightly influence handling stability for both tractor semi-trailers. Therefore, to ensure the driving safety of tractor semi-trailers when the tractor has only one nonsteering rear axle, much stricter restraints should be imposed on driving speed, and the loaded mass must not exceed the rated load of the trailer.

1. Introduction

Tractor semitrailers make up a large proportion of road transport vehicles due to their high transportation efficiency and economy. In 2010, the number of lorries in China reached 17 million, of which 8.7% were tractor semitrailers, and the number is growing at a rate of 20% per year.

At the same time, the number of traffic accidents involving tractor semitrailers remains high, and these accidents often cause serious personal injuries and property damage. Chinese statistics show that more than 1 in 3 tractors caused an accident on the highway in 2010, leaving 10,881 people dead.

Steering instability is one of the main contributors to traffic accidents involving tractor semitrailers. But, until now, little attention has been paid to it, and there are relatively few works on the topic [1–3]. This makes such a study essential.

Two different approaches are generally used to analyze vehicle handling stability: open and closed loop measurements. Unlike the former, the latter does not reflect the vehicle's inherent stability because driver behavior is taken into consideration [4–6]. As a result, most scholars prefer to use open loop analysis.

There are many methods of analyzing the open loop handling stability of vehicles, including (among others) dynamic theory [4–10] and graphical representations [11–13]. One of the most useful methods is the handling diagram invented by Pacejka [14–16], which has been used extensively due to its intuitive approach and convenience [11, 12, 17–22]. In [11, 12, 17, 18], handling diagrams were used to describe a general vehicle's handling stability based on linear or nonlinear models of vehicle dynamics. In [19–21], the handling stability of special vehicles, such as heavy trucks and multiwheel combat vehicles, was analyzed using handling diagrams. As a result, compared to dynamic theory, handling diagrams are easy to obtain and comprehensively describe handling stability for most vehicles. However, for vehicles with a locked differential or tandem rear axle, handling diagrams may be inadequate as the longitudinal forces on the two sides of the vehicle are not equal; see [22]. The new concept of the handling surface has been created to describe vehicle handling stability in these cases [23, 24].

To date, there has been too much focus on the handling stability of trucks and passenger cars, but few studies have attempted to characterize that of tractor semitrailers. Therefore, in this paper, the handling stability of tractor semitrailers is studied using handling diagrams, and the impact of vehicle design parameters and motion variables on the vehicle handling stability is analyzed. The research results have great importance for the development of active/passive safety control systems and for ensuring driving safety.

Given the importance of the number of nonsteering rear axles (on the tractor) on the handling stability of a tractor semitrailer and given that most tractors have one or two, tractor semitrailer when the tractor has two nonsteering rear axles and that when the tractor has one nonsteering rear axle are the two cases discussed in this paper. For simplicity, we refer to them as tractor semitrailers with two rear axles/one rear axle.

2. Handling Stability Analysis of the Tractor Semitrailer

First, we establish the handling equation to be used in plotting the handling diagram.

2.1. Review of Handling Equation for Simple Vehicles

In [14], Pacejka supposed that the steer angle is small, and the vehicle can be simplified to a bicycle model, and the handling equation for a simple vehicle can be expressed as

$$\delta - \frac{l}{R} = \alpha_f - \alpha_r = a_y \left(\frac{F_{zf}}{c_{\alpha f}} - \frac{F_{zr}}{c_{\alpha r}} \right), \quad (2.1)$$

where δ is the steering angle, l is the vehicle's wheelbase, and R is the turning radius. α_f and α_r are the side-slip angles of the front and rear tires, respectively. $c_{\alpha f}$ and $c_{\alpha r}$ are the cornering stiffness of the front and rear tires, respectively. F_{zf} and F_{zr} are the vertical loads of the front and rear tires, respectively. a_y is lateral acceleration.

The compliance factor K and the lateral acceleration are defined as follows:

$$K = \frac{F_{zf}}{c_{\alpha f}} - \frac{F_{zr}}{c_{\alpha r}}, \quad (2.2)$$

$$a_y = \frac{V^2}{gR'}, \quad (2.3)$$

where V is driving speed.

2.2. Handling Equations for Tractor Semitrailers

Based on the handling equation for simple vehicles, Winkler [19] produced different handling equations for complex vehicles, which may be presented in any of the following forms:

$$\delta - \frac{l}{R} = \alpha_f - \alpha_r = f_1(a_y, V), \quad (2.4)$$

$$\delta - \frac{l}{R} = \alpha_f - \alpha_r = f_2\left(a_y, \frac{1}{R}\right), \quad (2.5)$$

$$\delta - \frac{l}{R} = \alpha_f - \alpha_r = f_3\left(V, \frac{1}{R}\right). \quad (2.6)$$

Different from (2.1), except for being a function of lateral acceleration, the tire-side slip angle of complex vehicles may be related to the turning radius (when considering multiple nonsteering rear axles) or the driving speed (when considering the hysteresis effect of the tires). Equation (2.6), which is just a complement to (2.4), and (2.5), cannot be used individually.

It can be seen from (2.1), (2.4), and (2.5) that the vehicle handling stability is determined by the difference between the side-slip angles of the front and rear tires. Unlike a single-unit vehicle, a tractor semitrailer is composed of two vehicle units—a tractor and a trailer. The latter is attached to the former at an articulated point, using a fifth wheel coupling. Therefore, handling stability analysis of a tractor semitrailer is more complex as it must take the handling stabilities of each vehicle unit into account. For simplicity, we suppose that the tire-ground adhesion coefficient is big enough, and the trailer will not experience side-slip. Then, the tractor semitrailer's handling stability will be determined by the difference between the side-slip angles of the tractor's front and rear tires.

According to market surveys, most tractors have one or two nonsteering rear axles, those with two being most popular. Considering the impact of multiple nonsteering axles on vehicle handling stability, the following handling equation introduced by Winkler [19] can be used as the handling equation for a tractor semitrailer with two rear axles:

$$\delta - \frac{l}{R} = \frac{(1/IR)(1 + c_{\alpha r}/c_{\alpha f}) \sum_{i=1}^n \Delta i^2}{n + a_y(F_{zf}/c_{\alpha f} - F_{zr}/c_{\alpha r})}, \quad (2.7)$$

where Δi is the longitudinal distance from the compound rear axle of the tractor to the i th rear axle and the compound rear axle is located at the longitudinal position about which the vertical loads carried by all rear tires produce a net pitch moment of zero.

To simplify our analysis of the handling stability, we define the following quantity:

$$T = \frac{1}{2} \sum_{i=1}^2 \Delta i^2. \quad (2.8)$$

The nonlinear characteristics of tires on heavy vehicles mainly reflect the relationship between the tires' vertical loads and cornering stiffness, which is commonly described by the following equation introduced by Frenzo et al. [24]:

$$c_\alpha = c + c_1 F_z + c_2 F_z^2, \quad (2.9)$$

where c , c_1 , c_2 are constants.

Heavy vehicle tires have more stiffness, which makes the tire side-slip angle smaller and keeps the relation between the cornering force and the cornering stiffness fairly linear. This fact can be expressed by

$$F_y = \alpha c_\alpha, \quad (2.10)$$

where F_y is the tires' cornering force.

Now we define the following quantities:

$$\begin{aligned} C(a_y) &= \frac{c_{\alpha r}}{c_{\alpha f}}, \\ K(a_y) &= \frac{F_{zf}}{c_{\alpha f}} - \frac{F_{zr}}{c_{\alpha r}}. \end{aligned} \quad (2.11)$$

According to (2.7)–(2.11), the handling equation for a tractor semitrailer with two rear axles can be written as follows:

$$\delta - \frac{l}{R} = \frac{T}{lR} (1 + C(a_y)) + a_y K(a_y). \quad (2.12)$$

The total differentials of $\delta - l/R$ and a_y can be solved to analyze the change in vehicle handling stability with respect to lateral acceleration:

$$d\left(\delta - \frac{l}{R}\right) = \frac{\partial \delta}{\partial V} dV + \left(\frac{\partial \delta}{\partial(l/R)} - 1\right) d\left(\frac{l}{R}\right), \quad (2.13)$$

$$da_y = \frac{2V}{gR} dV + \frac{V^2}{g} d\left(\frac{l}{R}\right). \quad (2.14)$$

According to (2.12), the partial differentials of δ with respect to V and l/R are as follows:

$$\frac{\partial \delta}{\partial V} = \frac{2V}{gR} \left[K(a_y) + a_y \frac{dK(a_y)}{da_y} + \frac{dC(a_y)}{da_y} \right], \quad (2.15)$$

$$\frac{\partial \delta}{\partial(l/R)} - 1 = \frac{T}{l^2} (1 + C(a_y)) + \frac{V^2}{gl} \left[K(a_y) + a_y \frac{dK(a_y)}{da_y} + \frac{dC(a_y)}{da_y} \right]. \quad (2.16)$$

Supposing that the turning radius remains constant, we can substitute (2.15) and (2.14) into (2.13) to give

$$\left. \frac{d(\delta - l/R)}{da_y} \right|_{R=R_c} = K(a_y) + \frac{dC(a_y)}{da_y} + a_y \frac{dK(a_y)}{da_y}. \quad (2.17)$$

Now, supposing that the driving speed remains constant, we can substitute (2.16) and (2.14) into (2.13) to give

$$\left. \frac{d(\delta - l/R)}{da_y} \right|_{V=V_c} = \frac{Tg}{IV^2} (1 + C(ay)) + \left[K(a_y) + a_y \frac{dK(a_y)}{da_y} + \frac{dC(a_y)}{da_y} \right]. \quad (2.18)$$

Equations (2.17) and (2.18) give the changes in vehicle handling stability with respect to lateral acceleration holding the turning radius and driving speed constant, respectively.

The handling diagram that is obtained by holding the turning radius constant is called the R -handling diagram, and the corresponding handling curves are called R -handling curves. Similarly the V -handling diagram and V -handling curves refer to the case where the driving speed is held constant.

For a tractor semitrailer with one rear axle, the handling equation and partial differential with respect to a_y are as follows:

$$\begin{aligned} \delta - \frac{l}{R} &= a_y K(a_y), \\ \left. \frac{d(\delta - l/R)}{da_y} \right|_{V=V_c} &= \left. \frac{d(\delta - l/R)}{da_y} \right|_{R=R_c} = K(a_y) + a_y \frac{dK(a_y)}{da_y}. \end{aligned} \quad (2.19)$$

3. Solution of the Tractor Tires' Vertical Load

From Section 2.2, we know that the handling stability of a tractor semitrailer is actually determined by the vertical load distribution over the tractor's front and rear tires.

For analytical clearness, the tractor and its trailer are separated at the articulated point, and corresponding forces are applied to both vehicle units, as shown in Figure 1.

Under steady-state turning, the articulated angle θ between the tractor and trailer always remains small. Thus, we can say

$$\cos \theta \approx 1, \quad \sin \theta \approx 0. \quad (3.1)$$

The semitrailer is partly supported by the tractor once the two are connected. The load that comes from the trailer and is carried by the tractor can be expressed by

$$F = \frac{b'}{l'} (m_{s2} + m_c) g, \quad (3.2)$$

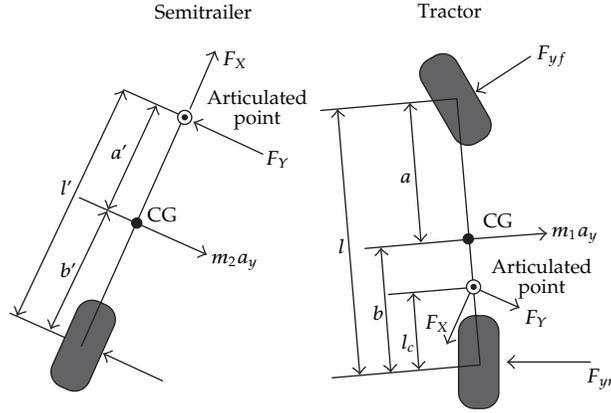


Figure 1: Representation of tractor semitrailer: main dimensions and lateral forces.

where b' is the distance from the center of the trailer's sprung mass to the compound rear axle, l' is the distance from the center of the kingpin to the compound rear axle, m_{s2} is the trailer's sprung mass, and m_c is the loaded mass.

The fifth wheel is installed on the rear of the tractor, which moves the center of gravity (CG) of the tractor's sprung mass backwards when it is towing a trailer. The distance the CG moves can be expressed by

$$\Delta l = \frac{(b - l_c)b'(m_{s2} + m_c)}{l'(m_{sf1} + m_{sr1}) + b'(m_{s2} + m_c)}, \quad (3.3)$$

where b is the distance from the CG of the tractor's sprung mass to its compound rear axle, l_c is the fifth wheel lead, which is the distance from the center of the fifth wheel's jaw to the CG of the tractor's sprung mass, and m_{sf1} and m_{sr1} are the sprung masses of the tractor's front and rear axles, respectively.

According to (3.2) and (3.3), the vertical loads of the tractor's front and rear axles when the tractor is towing a trailer can be presented as follows:

$$\begin{aligned} F_{zf} &= \left[(m_{sf1} + m_{uf1}) + \frac{l_c b'}{l l'} (m_{s2} + m_c) \right] g, \\ F_{zr} &= \frac{a}{l} (m_{sr1} + m_{ur1}) g + \frac{l - l_c b'}{l l'} (m_{s2} + m_c) g, \end{aligned} \quad (3.4)$$

where m_{uf1} and m_{ur1} are the unsprung masses of the tractor's front and rear axles, respectively, and a is the distance from the CG of the tractor's sprung mass to the front axle.

Side-to-side lateral load transfer is generated when a lateral force acts on the vehicles. As seen from Figure 2, the lateral load transfers of the tractor's front and rear tires can be expressed by

$$\begin{aligned} \Delta F_{zf} &= \frac{a_y m_s g}{d_f} \left[\frac{h_s}{1 + k_r/k_f - m_s g h_s/k_f} + \frac{b - \Delta l}{l} h_f \right], \\ \Delta F_{zr} &= \frac{a_y m_s g}{d_r} \left[\frac{h_s}{1 + k_f/k_r - m_s g h_s/k_r} + \frac{a + \Delta l}{l} h_r \right], \end{aligned} \quad (3.5)$$

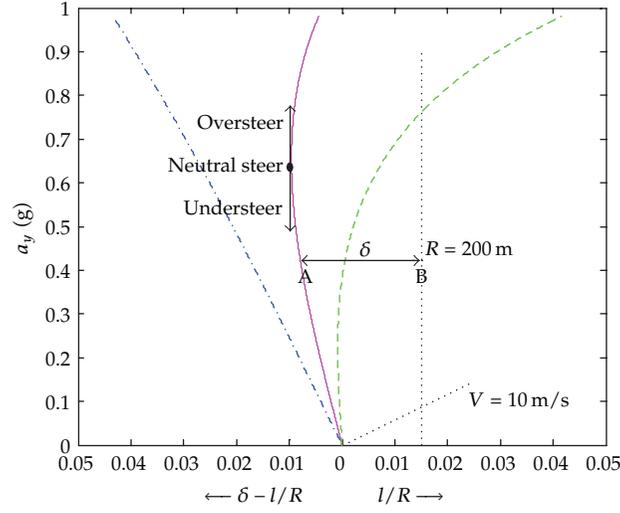


Figure 3: Handling diagram for a passenger car.

Table 1: Equivalent definitions for understeer, neutral steer, and oversteer of a complex vehicle.

At a constant radius	Oversteer	$d(\delta - l/R)/da_y _{R=R_c} < 0$	$\partial\delta/\partial V < 0$
	Neutral steer	$d(\delta - l/R)/da_y _{R=R_c} = 0$	$\partial\delta/\partial V = 0$
	Understeer	$d(\delta - l/R)/da_y _{R=R_c} > 0$	$\partial\delta/\partial V > 0$
At a constant velocity	Oversteer	$d(\delta - l/R)/da_y _{V=V_c} < 0$	$\partial\delta/\partial(l/R) < 0$
	Neutral steer	$d(\delta - l/R)/da_y _{V=V_c} = 0$	$\partial\delta/\partial(l/R) = 0$
	Understeer	$d(\delta - l/R)/da_y _{V=V_c} > 0$	$\partial\delta/\partial(l/R) > 0$

The load distribution across the tractor's front and rear axles when it is towing a trailer is determined by the fifth wheel lead, which can be adjusted within a certain range to allow the fifth wheel to engage with the kingpin. The fifth wheel for the tractor with two rear axles is arranged between the two rear axles, and the fifth wheel lead is about 0.3 m. According to collected statistics, the distance between the two rear axles is generally 1.45 m, 1.4 m, 1.37 m, 1.35 m, 1.3 m, and so forth, and 1.35 m is the most common distance. Therefore, the adjustment range for the fifth wheel lead for a tractor with two rear axles is somewhere between 0 and 0.675 m.

To analyze the influence of the fifth wheel lead on the handling stability of a tractor semitrailer with one rear axle, we assume that the adjustment range is the same as for the tractor with two rear axles.

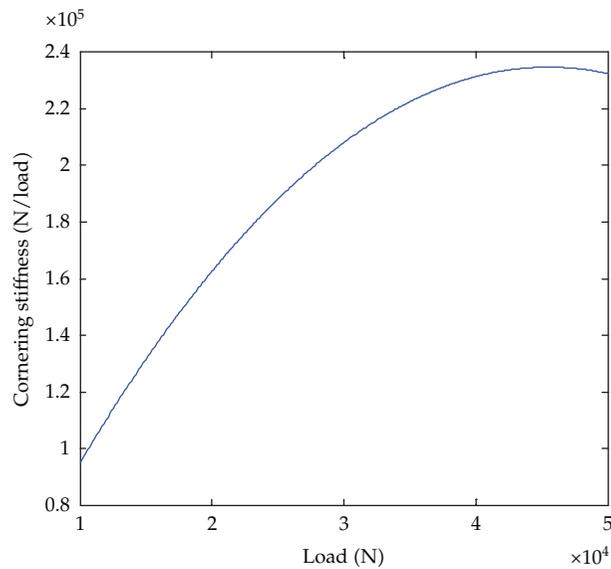
All rear tires are assumed to be the same and to support equal loads. Taking lateral load transfer into consideration, tire cornering stiffness can be calculated as the average of the right and left tires' cornering stiffness, and the vertical load can also be assumed to be the average of the right and left loads.

The constants in (2.9) are assumed to take the following values (see in [25]):

$$\begin{aligned}
 c &= 5000 \text{ N/rad}, \\
 c_1 &= 10.1/\text{rad}, \\
 c_2 &= -11.1/(\text{N} \cdot \text{rad}).
 \end{aligned} \tag{4.1}$$

Table 2: Characteristics of the tractor semi-trailer with two rear axles.

Parameter	Value	Parameter	Value
m_{sf1}	5069 kg	m_{sr1}	2431 kg
m_{uf1}	790 kg	m_{ur1}	1590 kg
a	1.288 m	b	2.687 m
k_f	140000 N/rad	l_c	0.3 m
k_r	202000 N/rad	d_f	2.2 m
d_r	1.8 m	h_f	0.5 m
h_r	0.5 m	h_s	0.242 m
m_{s2}	6000 kg	a'	4.77 m
b'	3.34 m	l	3.975 m
l'	8.11 m	T	0.36 m

**Figure 4:** Tire cornering stiffness as a function of vertical load.

Based on the vehicle characteristics presented in Table 2 and the tire property parameters, we can calculate curves showing tire cornering stiffness and the ratio of rear tires' to front tires' cornering stiffness. These are presented in Figures 4 and 5.

4.3. Influence of Vehicle Parameters on Handling Stability

We use Matlab to simulate the influence of different vehicle parameters on the handling stability of the tractor semitrailer with one or two rear axles.

4.3.1. Tractor Semitrailer with Two Rear Axles

We use (2.12) to simulate the vehicle handling stability of a tractor semitrailer with two rear axles under either a constant driving speed or turning radius. The simulation results are presented in Figures 6 and 7.

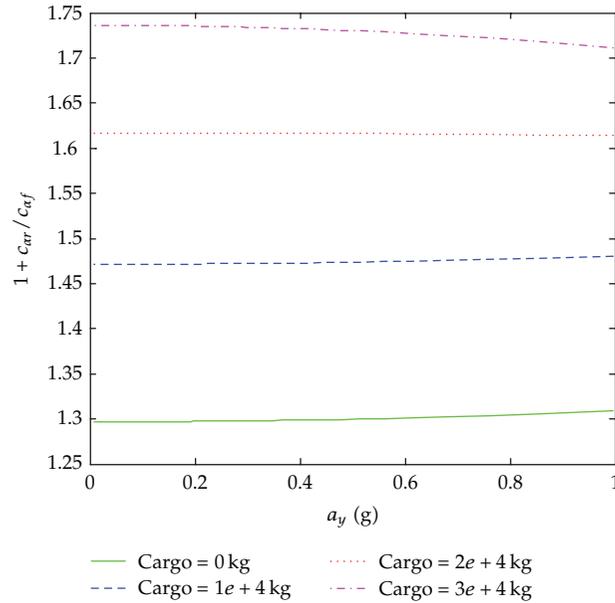


Figure 5: Ratio of rear tire's to front tire's cornering stiffness.

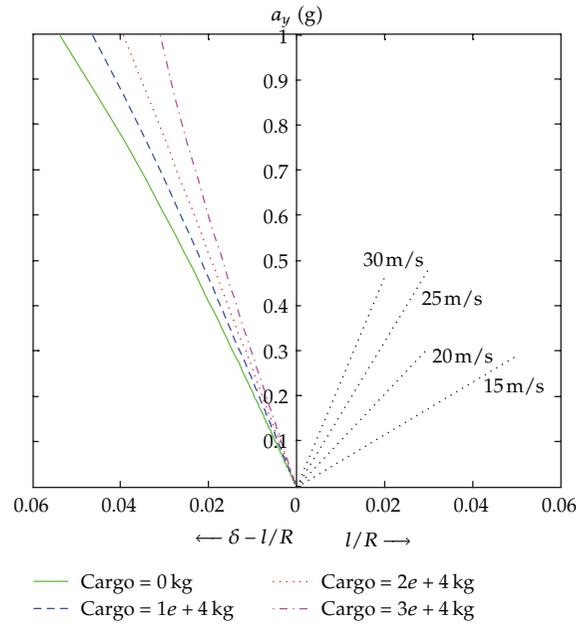


Figure 6: V-handling diagram for tractor semitrailer with two rear axles.

As can be seen in Figures 6 and 7, the V-handling curves and R-handling curves for a tractor semitrailer with two rear axles are quite similar to one another as the other parameters vary. Figure 5 shows that the range of $1 + c_{ar}/c_{af}$ is 1.3–1.75. As T/lR is quite small for most tractors, $T/lR(1 + C(a_y))$, the multiple nonsteering rear axles term in (2.12) has quite a small impact on the vehicle handling stability. However, a slight influence can still be seen in the

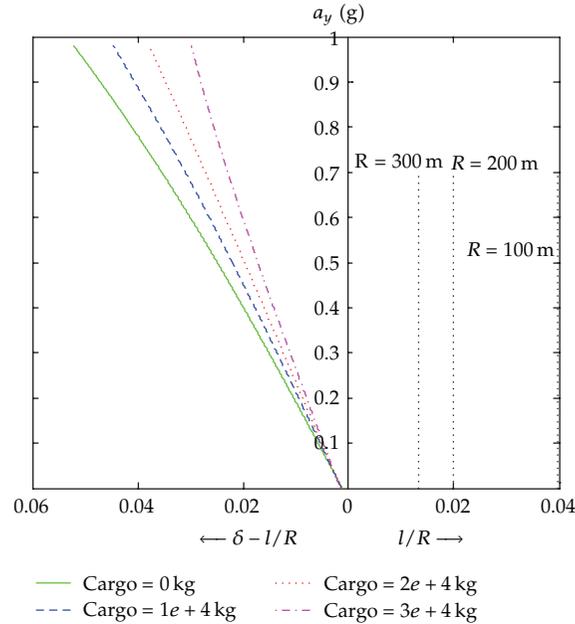


Figure 7: *R*-handling diagram for tractor semitrailer with two rear axles.

R-handling diagram, as the starting points of the *R*-handling curves are not at the origin but at a positive value in the direction of the $\delta - l/R$ axis. The same explanation can also be obtained from (2.17) and (2.18).

The most important finding that can be drawn from Figures 6 and 7 is that a tractor semitrailer with two rear axles has great handling stability within the trailer’s rated load range. The tractor has four times as many rear tires as front tires, making both the vertical load and the load transfer of a single rear tire smaller than those of a single front tire. In addition to the rear tire’s strong stiffness, $K(a_y)$ in (2.12) is positive, and the vehicle has great handling stability. However, there is a fall in the value of $dC(a_y)/da_y$ in (2.17) due to the increase in loaded mass, and this becomes negative when the loaded mass reaches a certain value. As the minimum value $dC(a_y)/da_y$ can reach is -0.06 , which in absolute value terms is larger than $K(a_y)$ ($0.028-0.0525$), the vehicle handling stability becomes poor under an increase in the loaded mass.

Therefore, the excellent handling stability of a tractor semitrailer with two rear axles owes more to its rear tires than to its multiple nonsteering rear axles. Unlike the conclusions drawn in [19], we find that multiple nonsteering rear axles do not strengthen the vehicle’s handling stability but weaken it.

A handling diagram for a tractor semitrailer with two rear axles and varying fifth wheel leads is presented in Figure 8. The influence of the fifth wheel lead on vehicle handling stability is quite small. However, the larger the fifth wheel lead is, the better handling stability the vehicle has.

4.3.2. Tractor Semitrailer with One Rear Axle

For a tractor semitrailer with one rear axle, the *V*-handling and *R*-handling curves coincide. However, for analytical convenience, we present the two diagrams separately, in Figures 9 and 10.

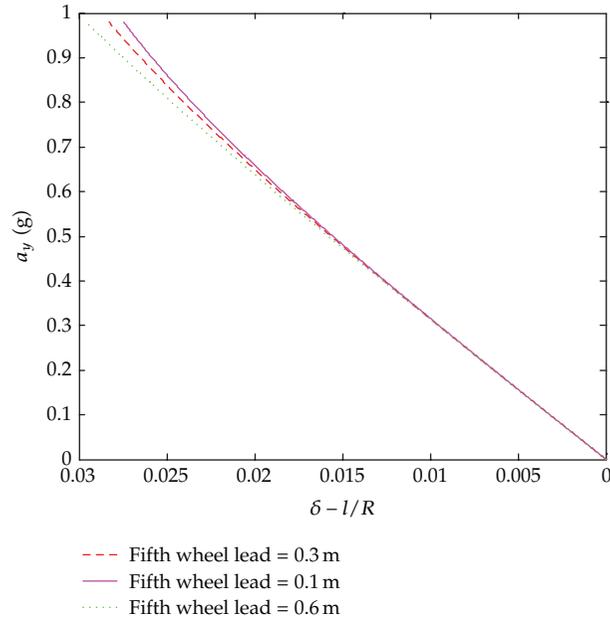


Figure 8: Handling diagram for a tractor semitrailer with two rear axles and varying fifth wheel leads.

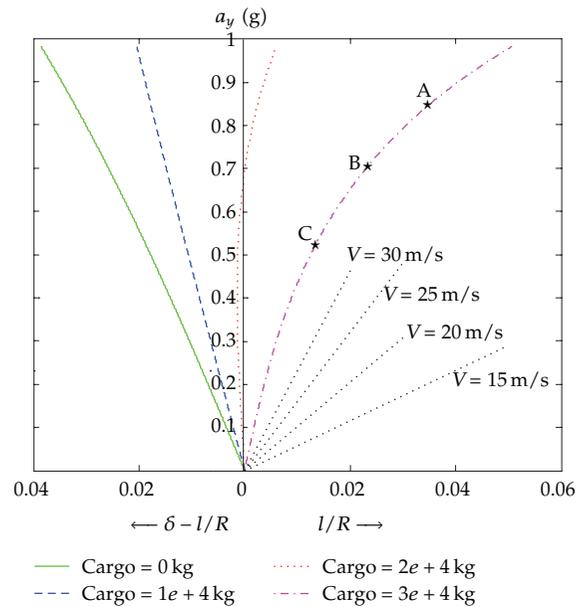


Figure 9: V-handling diagram for tractor semitrailer with one rear axle.

As Figures 9 and 10 show, vehicle handling stability quickly becomes poor as the loaded mass increases. In a fully loaded situation, the vehicle handling stability presents an understeering characteristic in a narrow lateral acceleration range and then turns into a neutral steer or oversteering situation.

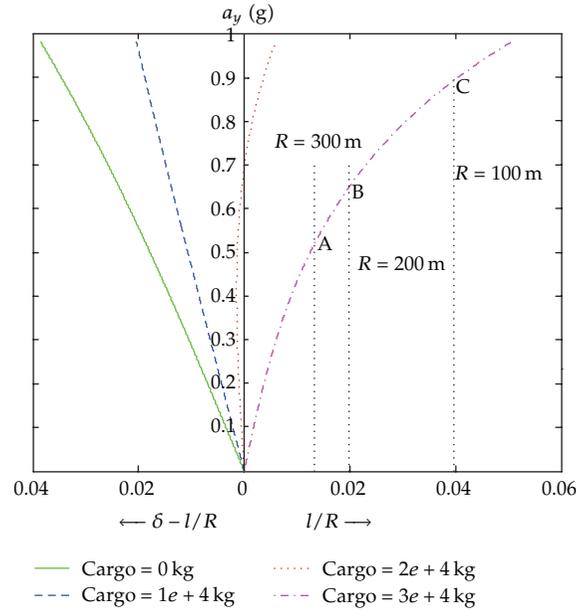


Figure 10: *R*-handling diagram for tractor semitrailer with one rear axle.

Figure 9 shows that the driving speed has a large impact on the vehicle handling stability when the loaded mass is large. When the loaded mass is 30,000 kg and driving speeds are 20 m/s, 25 m/s and 30 m/s, respectively, the critical lateral acceleration values at which instability may occur are at points A (0.8446 g), B (0.7031 g), and C (0.5203 g). An increase in driving speed will further weaken the vehicle handling stability and eventually result in instability. Considering heavy vehicles’ rollover thresholds, such vehicles are prone to rollover accidents at low speeds and handling instability accidents at high speeds, even if the turning radius is quite large ($R = 177$ m at point C).

Figure 10 shows that, when the loaded mass is 30,000 kg and the turning radii are 100 m, 200 m, and 300 m, respectively, the intersection points between the l/R lines and the *R*-handling curve occur at points C, B, and A. The physical meaning of these intersection points is that the vehicle loses steering capacity ($\delta = 0$) and rotates on its axis. The corresponding driving speeds for points A, B, and C are 39 m/s, 35.67 m/s, and 29.6 m/s, respectively, which are obtained from (2.3). Taking into account the strong impact of driving speed on vehicle handling stability, even if the turning radius is large and there is no apparent road curvature, the vehicle will still lose its original stability in these conditions.

Based on the previous analysis, a vehicle handling stability threshold defined by driving speed is more appropriate than one defined by lateral acceleration. This implies that it is important to set limits on the driving speeds of heavy vehicles on highways.

The handling diagram for a tractor semitrailer with one rear axle and varying fifth wheel leads is presented in Figure 11. The same conclusions can be drawn as in the case of two rear axles. Due to the poor handling stability of the tractor with one rear axle, the fifth wheel lead should be as large as possible when the loaded mass is large.

4.4. Conclusions

In this paper, we analyze the handling stability of tractor semitrailers with one or two rear axles, based on handling diagrams. The following conclusions are obtained.

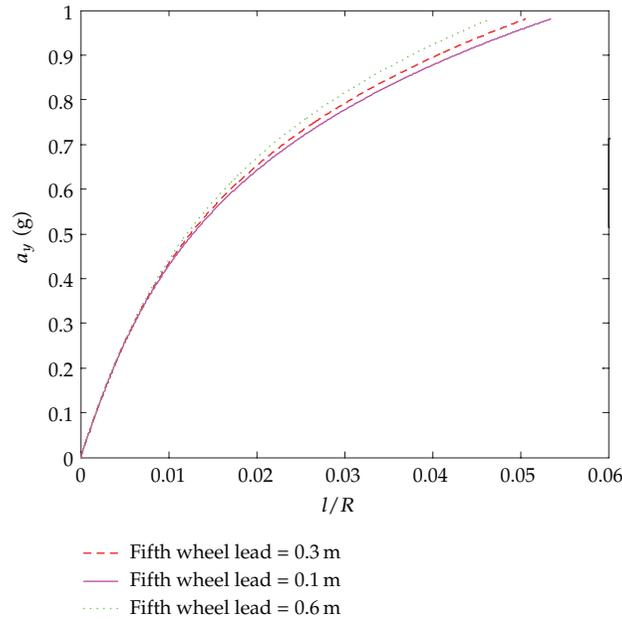


Figure 11: Handling diagram for a tractor semitrailer with one rear axle and varying fifth wheel leads.

- (1) Tractor semitrailers with two rear axles have better handling stability than those with just one. Within the rated load range, the former always show good steering performance. This result is due to the larger number of rear tires on the two-rear-axle vehicle.
- (2) For a tractor with one rear axle, towing a trailer with a suitable total mass and driving at a low speed are quite important. This tractor semitrailer will lose handling stability when overloaded or if driven at high speeds. The vehicle may even lose steering capacity at high speeds.
- (3) Driving speed has much more influence on vehicle handling stability than the turning radius. Even on a gentle curve, these vehicles should not be driven at speeds above a certain level.
- (4) Fifth wheel lead has a slight influence on the handling stability of tractor semitrailers with one and with two rear axles. However, for tractors with one rear axle, it is especially important that the fifth wheel lead should be as large as possible in order to improve vehicle handling stability.

Since we make the assumption in deriving the handling equations that the tire-ground adhesion coefficient is sufficiently large, the handling equations are limited to analyzing handling stability for vehicles driven under certain road and driving conditions only. Thus, handling stability analysis for tractor semitrailers driven on slippery roads will be conducted in a future study.

Acknowledgments

This research is supported by “Research on Evaluation and Detection Technology of Handling and Driving Stability for Commercial Vehicles (no. 2009BAG13A04)” as part of the Major Project of the 11th Five-Year National Science and Technology Support Program:

serious road traffic accident prevention and integrated disposal technology development and demonstration applications, organized by the Ministry of Science and Technology, Public Security, and Transportation of China.

References

- [1] H. G. Xu, T. Peng, and H. F. Liu, "Feedback linearization for steering stability control of tractor-semitrailer," *Journal of Jilin University*, vol. 42, no. 2, pp. 272–278, 2012.
- [2] Z. W. Guan, X. J. Wang, and M. F. Zheng, "Analysis of steering stability of tractor-semitrailer combination vehicle," *Automobile Technology*, no. 6, pp. 41–45, 2011.
- [3] S. Chang, H. G. Xu, and H. F. Liu, "Simulation on steering stability of 4WS tractor semi-trailer," in *Proceedings of the International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '09)*, pp. 355–358, April 2009.
- [4] W. H. Wang, W. Zhang, H. W. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research Part C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [5] W. H. Wang, "A digital-driving system for smart vehicles," *IEEE Intelligent Systems*, vol. 17, no. 5, pp. 81–83, 2002.
- [6] W. H. Wang, M. Yan, 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.
- [7] K. Hussain, W. Stein, and A. J. Day, "Modelling commercial vehicle handling and rolling stability," *Journal of Multi-Body Dynamics*, vol. 219, no. 4, pp. 357–369, 2005.
- [8] E. Dahlberg and N.G. Vagstedt, "Advantages of a simple approach modeling heavy vehicle handling," *SAE Special Publications*, vol. 1308, pp. 113–120, 1997.
- [9] W. Chen, Z. Zhang, and C. Zhou, "Simulation for the handling and stability of four-wheel steering vehicle based on Matlab/simulink," in *Proceedings of the 2nd International Conference on Transportation Engineering (ICTE '09)*, pp. 1908–1913, July 2009.
- [10] Q.F. Zhao, T. He, and W. J. Xu, "The research of vehicle handling stability based on ADAMS," *Applied Mechanics and Materials*, vol. 127, pp. 248–251, 2012.
- [11] K. Mathijs, *Longitudinal force distribution and road vehicle handling [Ph.D. thesis]*, Department of Applied Mechanics, Chalmers University of Technology, 2010.
- [12] N. Vincent, *Vehicle handling, stability, and bifurcation analysis for nonlinear vehicle models [M.S. thesis]*, Department of Mechanical Engineering, University of Maryland, 2005.
- [13] L. Liu, *Nonlinear analysis and control strategy evaluation on the stability of vehicle 3-DOF planar motion [Ph.D. thesis]*, College of Traffic, Jilin University, 2010.
- [14] H. B. Pacejka, "Simplified analysis of steady-state turning behaviour of motor vehicles—part 1. Handling diagrams of simple systems," *Vehicle System Dynamics*, vol. 2, no. 3, pp. 161–172, 1973.
- [15] H. B. Pacejka, "Simplified analysis of steady-state turning behaviour of motor vehicles—part 2: stability of the steady-state turn," *Vehicle System Dynamics*, vol. 2, no. 4, pp. 173–183, 1973.
- [16] H. B. Pacejka, "Simplified analysis of steady-state turning behaviour of motor vehicles—part 3: more elaborate systems," *Vehicle System Dynamics*, vol. 2, no. 4, pp. 185–204, 1973.
- [17] A. F. Daniel and S. Adam, "Passenger vehicle steady-state directional stability analysis utilizing EDVSM and SIMON," in *Proceedings of the WP HVE Forum*, San Francisco, Calif, USA, May 2004.
- [18] S. Lukowski, M. Momot, D. Kraemer, and D. Kunz, "Basic linear theory of handling and stability of automobiles," *Journal of Automobile Engineering*, vol. 223, no. 1, pp. 1–10, 2009.
- [19] C. B. Winkler, "Simplified analysis of the steady-state turning of complex vehicles," *Vehicle System Dynamics*, vol. 29, no. 3, pp. 141–180, 1998.
- [20] E. C. Yeh and Y. L. Chen, "Handling analysis of a motorcycle with added cambering of the front frame," *Vehicle System Dynamics*, vol. 19, no. 2, pp. 49–70, 1990.
- [21] M. J. Hillegass, J. G. Faller, M. S. Bounds, M. El-Gindy, and A. S. Joshi, "Validating the directional performance of multi-wheeled combat vehicle computer simulation models," in *Proceedings of the ASME International Mechanical Engineering Congress and Exposition (IMECE '04)*, pp. 781–789, November 2004.

- [22] F. Frendo, G. Greco, and M. Guiggiani, "Critical review of handling diagram and understeer gradient for vehicles with locked differential," *Vehicle System Dynamics*, vol. 44, no. 6, pp. 431–447, 2006.
- [23] F. Frendo, G. Greco, M. Guiggiani, and A. Sponziello, "The handling surface: a new perspective in vehicle dynamics," *Vehicle System Dynamics*, vol. 45, no. 11, pp. 1001–1016, 2007.
- [24] F. Frendo, G. Greco, M. Guiggiani, and A. Sponziello, "Evaluation of the vehicle handling performances by a new approach," *Vehicle System Dynamics*, vol. 46, no. 1, pp. 857–868, 2008.
- [25] D. J. M. Sampson, *Active roll control of articulated heavy vehicles [Ph.D. thesis]*, Churchill College, University of Cambridge, 2000.