

Journal of Advanced Transportation

Models and Technologies for Transport System Flow Analysis

Lead Guest Editor: Zhiyuan Liu

Guest Editors: Lele Zhang and David Z. W. Wang





Models and Technologies for Transport System Flow Analysis

Journal of Advanced Transportation

Models and Technologies for Transport System Flow Analysis

Lead Guest Editor: Zhiyuan Liu

Guest Editors: Lele Zhang and David Z. W. Wang



Copyright © 2019 Hindawi. All rights reserved.

This is a special issue published in "Journal of Advanced Transportation." 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

Francesco Bella, Italy
Abdelaziz Bensrhair, France
Cesar Briso-Rodriguez, Spain
María Calderon, Spain
Juan C. Cano, Spain
Giulio E. Cantarella, Italy
Maria Castro, Spain
Oded Cats, Netherlands
Anthony Chen, USA
Nicolas Chiabaut, France
Steven I. Chien, USA
Antonio Comi, Italy
Emanuele Crisostomi, Italy
Luca D'Acierno, Italy
Andrea D'Ariano, Italy
Alexandre De Barros, Canada
Stefano de Luca, Italy
Rocío de Oña, Spain
Luigi Dell'Olio, Spain
Cédric Demonceaux, France
Sunder Lall Dhingra, India
Vinayak Dixit, Australia
Yuchuan Du, China
Nour-Eddin El-faouzi, France

Juan-Antonio Escareno, France
David F. Llorca, Spain
Peter Furth, USA
Francesco Galante, Italy
Md. Mazharul Haque, Australia
Jérôme Haïri, France
Samiul Hasan, USA
Serge Hoogendoorn, Netherlands
Hocine Imine, France
Lina Kattan, Canada
Victor L. Knoop, Netherlands
Alain Lambert, France
Ludovic Leclercq, France
Jaeyoung Lee, USA
Seungjae Lee, Republic of Korea
Zhi-Chun Li, China
Yue Liu, USA
Jose R. Martinez-De-Dios, Spain
Filomena Mauriello, Italy
Monica Menendez, UAE
Rakesh Mishra, UK
Andrea Monteriù, Italy
Giuseppe Musolino, Italy
Jose E. Naranjo, Spain

Aboelmaged Noureldin, Canada
Eneko Osaba, Spain
Eleonora Papadimitriou, Netherlands
Dongjoo Park, Republic of Korea
Paola Pellegrini, France
Luca Pugi, Italy
Nandana Rajatheva, Finland
Hesham Rakha, USA
Prianka N. Seneviratne, Philippines
Fulvio Simonelli, Italy
Richard S. Tay, Australia
Martin Trépanier, Canada
Pascal Vasseur, France
Antonino Vitetta, Italy
Francesco Viti, Luxembourg
S. Travis Waller, Australia
Yair Wiseman, Israel
George Yannis, Greece
Shamsunnahar Yasmin, USA
Jacek Zak, Poland
Guohui Zhang, USA
Ning Zhang, USA
G. Homem de Almeida Correia, Netherlands

Contents

Models and Technologies for Transport System Flow Analysis

Zhiyuan Liu , Lele Zhang, and David Z. W. Wang 

Editorial (4 pages), Article ID 6964943, Volume 2019 (2019)

On the Effects of Various Measures of Performance Selections on Simulation Model Calibration Performance

Chen Wang  and Chengcheng Xu 

Research Article (16 pages), Article ID 3839814, Volume 2018 (2019)

Analysis on Port and Maritime Transport System Researches

Lingshan Chen, Xiang Xu , Ping Zhang , and Xiaotian Zhang

Research Article (20 pages), Article ID 6471625, Volume 2018 (2019)

Application of Finite Mixture of Logistic Regression for Heterogeneous Merging Behavior Analysis

Gen Li 

Research Article (9 pages), Article ID 1436521, Volume 2018 (2019)

Characterizing Critical Transition State for Network Fundamental Diagram

Rongrong Hong, Chengchuan An, Zhenbo Lu , Jingxin Xia , Qinghui Nie, and Wenming Rao

Research Article (13 pages), Article ID 4839729, Volume 2018 (2019)

Empirical Approximation for the Stochastic Fundamental Diagram of Traffic Flow on Signalized Intersection

Nan Zhang , Xiaoguang Yang , and Wangjing Ma 

Research Article (17 pages), Article ID 4603614, Volume 2018 (2019)

The Impact of Aggressive Driving Behavior on Driver-Injury Severity at Highway-Rail Grade Crossings Accidents

Changxi Ma , Wei Hao , Wang Xiang, and Wei Yan

Research Article (10 pages), Article ID 9841498, Volume 2018 (2019)

Driving Risk Detection Model of Deceleration Zone in Expressway Based on Generalized Regression Neural Network

Weiwei Qi , Zhexuan Wang, Ruru Tang, and Linhong Wang 

Research Article (8 pages), Article ID 8014385, Volume 2018 (2019)

Modelling Rail-Based Park and Ride with Environmental Constraints in a Multimodal Transport Network

Xinyuan Chen and Inhi Kim 

Research Article (15 pages), Article ID 2310905, Volume 2018 (2019)

Predicting and Visualizing the Uncertainty Propagations in Traffic Assignments Model Using Monte Carlo Simulation Method

Mundher Seger  and Lajos Kisgyörgy

Research Article (11 pages), Article ID 9825327, Volume 2018 (2019)

Enhancing Freeway Safety through Intervening in Traffic Flow Dynamics Based on Variable Speed Limit Control

Jinming You, Shouen Fang, Lanfang Zhang, John Taplin, and Jingqiu Guo 
Research Article (10 pages), Article ID 3610541, Volume 2018 (2019)

Bike-Sharing Static Rebalancing by Considering the Collection of Bicycles in Need of Repair

Sheng Zhang , Guanhua Xiang, and Zhongxiang Huang 
Research Article (18 pages), Article ID 8086378, Volume 2018 (2019)

Impact of a New Metro Line: Analysis of Metro Passenger Flow and Travel Time Based on Smart Card Data

Xiao Fu  and Yu Gu 
Research Article (13 pages), Article ID 9247102, Volume 2018 (2019)

Improving Traffic State Prediction Model for Variable Speed Limit Control by Introducing Stochastic Supply and Demand

Yuwei Bie , Mudasser Seraj, Can Zhang, and Tony Z. Qiu 
Research Article (12 pages), Article ID 7959815, Volume 2018 (2019)

A Two-Layer Network Dynamic Congestion Pricing Based on Macroscopic Fundamental Diagram

Bangyang Wei  and Daniel(Jian) Sun 
Research Article (11 pages), Article ID 8616120, Volume 2018 (2019)

Automatic Estimation Method for Intersection Saturation Flow Rate Based on Video Detector Data

Linhong Wang , Yunhao Wang, and Yiming Bie 
Research Article (9 pages), Article ID 8353084, Volume 2018 (2019)

Optimal Bus-Bridging Service under a Metro Station Disruption

Haodong Yin, Jianjun Wu , Huijun Sun , Yunchao Qu, Xin Yang, and Bo Wang
Research Article (16 pages), Article ID 2758652, Volume 2018 (2019)

A Crash Surrogate Metric considering Traffic Flow Dynamics in a Motorway Corridor

Xu Wang and Kai Liu 
Research Article (7 pages), Article ID 9349418, Volume 2018 (2019)

Optimal Design of Transportation Networks with Automated Vehicle Links and Congestion Pricing

Yipeng Ye  and Hua Wang 
Research Article (12 pages), Article ID 3435720, Volume 2018 (2019)

Minimizing the Average Delay at Intersections via Presignals and Speed Control

Mina Ghanbarikarekani , Xiaobo Qu , Michelle Zeibots, and Weiwei Qi 
Research Article (8 pages), Article ID 4121582, Volume 2018 (2019)

An Association Rule Based Method to Integrate Metro-Public Bicycle Smart Card Data for Trip Chain Analysis

De Zhao , Wei Wang, Ghim Ping Ong, and Yanjie Ji 
Research Article (11 pages), Article ID 4047682, Volume 2018 (2019)

Editorial

Models and Technologies for Transport System Flow Analysis

Zhiyuan Liu ¹, Lele Zhang,² and David Z. W. Wang ³

¹*Jiangsu Key Laboratory of Urban ITS, Jiangsu Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, China*

²*School of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia*

³*School of Civil and Environmental Engineering, Nanyang Technological University, Singapore*

Correspondence should be addressed to Zhiyuan Liu; zhiyuanl@seu.edu.cn

Received 28 November 2018; Accepted 28 November 2018; Published 10 January 2019

Copyright © 2019 Zhiyuan Liu 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.

Flow analysis is the core of many well-recognized technologies for the evaluation and improvement of transport systems, including the analytical methods and simulation software packages for transport system planning and design, traffic control, traffic safety analysis, and demand management. Urban and regional transport systems belong to a highly complex sector that involves the intricate activities of millions of commuters each day, which makes it highly challenging to accurately analyze and predict the flows. Herein, these flows also cover the dynamics in different subsystems, including the urban road networks and public transport systems; thus how to reasonably address the interactions and correlations between these subsystems is still a difficult and prevailing problem in the transport engineering and science area. In addition, new emerging sharing mobilities (such as ride-sharing, car sharing, and bike sharing) have provided new dimensions to travelers for the trip choices, which have introduced new challenges to the studies of transport system flows. To this end, this special issue aims to address the cutting-edge ideas, knowledge, methodologies, techniques, and practices in the broad areas of transport system flow analysis. There were 70 papers submitted to this special issue, 20 of which were accepted for publication. As the guest editors of this special issue, we would like to summarize the accepted papers as follows.

“Analysis on Port and Maritime Transport System Researches” by L. Chen et al.

Container port and maritime transportation has grown as a unique academic field in recent decades. In this study, rigorous bibliometric analysis methods are employed to review the

literature of container port and maritime logistics research to illustrate the evolution of this field. A four-step method is designed to collect and process data. The authors rank the most active researchers, affiliations, and countries/regions adopting two scoring criteria and identify overall research terms at the microlevel by the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm, which provides a better understanding of how maritime transportation research has been undertaken in a quantitative manner.

“Application of Finite Mixture of Logistic Regression for Heterogeneous Merging Behavior Analysis” by G. Li

In this research, a two-component finite mixture of logistic regression model is applied to analyze the vehicle trajectory data collected on a highway segment and discovered two major merging behaviors of drivers: risk-rejecting and risk-taking.

“Characterizing Critical Transition State for Network Fundamental Diagram” by R. Hong et al.

This paper proposes a clustering method based on the Gaussian mixture model to classify the critical transition state (CTS) of traffic flow using the time series data of network aggregated density and flow. The findings of this research are helpful to understand the temporal evolution process of traffic flow.

“Empirical Approximation for the Stochastic Fundamental Diagram of Traffic Flow on Signalized Intersection” by N. Zhang et al.

In this study, a stochastic traffic flow model is designed, and a shockwave speed based stochastic fundamental diagram is studied for a road segment between two signalized intersections for capturing the density-flow data scattering. The distribution of shockwave speeds is obtained by the variational Bayesian learning method from the field data.

“The Impact of Aggressive Driving Behavior on Driver-Injury Severity at Highway-Rail Grade Crossings Accidents” by C. Ma et al.

In this research, the authors analyze the accident data at US highway-rail grade crossings to explore the determinants of driver-injury severity with and without aggressive driving behaviors. A mixed logit modelling approach with a likelihood ratio test is utilized to estimate the problem. Results show that young male drivers, bad weather, peak-hour conditions, and driving in open space areas were more likely to express aggressive driving behaviors outwardly with high level injury severity given a highway-rail grade crossing accident happened.

“Driving Risk Detection Model of Deceleration Zone in Expressway Based on Generalized Regression Neural Network” by W. Qi et al.

Drivers' mistakes may cause some traffic accidents, and such accidents can be avoided if prompt advices could be given to drivers. So, how to detect driving risk is the key factor. Firstly, the selected parameters of vehicle movement are reaction time, acceleration, initial speed, final speed, and velocity difference. The ANOVA results show that the velocity difference is not significant on different driving state, and the other four parameters can be used as input variables of neural network models in deceleration zone of expressway, which have fifteen different combinations. Then, the detection model results indicate that the prediction accuracy rate of testing set is up to 86.4%. An interesting finding is that the number of input variables is positively correlated with the prediction accuracy rate.

“Modelling Rail-Based Park and Ride with Environmental Constraints in a Multimodal Transport Network” by X. Chen and I. Kim

This paper proposes a rail based Park-and-Ride (RPR) scheme to promote public transport in the multimodal transport network. To remedy the heterogeneous distribution of vehicle pollutants in the network, regulations in environmental sensitive districts are required and studied in this paper. To quantitatively evaluate and analyze this joint RPR and environmental regulation strategy in multimodal transport systems, this paper develops an environmental constrained combined modal split and traffic assignment (EC-CMSTA) model. The proposed formulation adopts the concept of fix-point to reformulate the nonlinear complementarity conditions associated with the combined modal split and user equilibrium conditions, which is subsequently incorporated into a VI formulated nonlinear complementarity conditions associated with environmental constraints. A customized projection based self-adaptive gradient projection (SAGP) algorithm is then developed to solve the problem.

“Predicting and Visualizing the Uncertainty Propagations in Traffic Assignments Model Using Monte Carlo Simulation Method” by M. Seger and L. Kisgyörgy

In this research, the authors develop a five stages model based on Monte Carlo (MC) simulation to predict and visualize traffic flow and its uncertainty in traffic assignment models. After generating origin-destination (OD) matrices using MC simulations and simulating the traffic assignment with VISUM, the predicted traffic flow on each link is categorized based on the bias and variability.

“Enhancing Freeway Safety through Intervening in Traffic Flow Dynamics Based on Variable Speed Limit Control” by J. You et al.

This paper proposes an innovative Variable Speed Limit (VSL) based approach to manage crash risks by intervening in traffic flow dynamics on freeways using High Definition Monitoring System (HDMS) data. A binary logistic regression model is built to estimate crash risk. Microsimulations have been conducted to verify the proposed method with the AIMSUN simulation software and some insights are obtained.

“Bike-Sharing Static Rebalancing by Considering the Collection of Bicycles in Need of Repair” by S. Zhang et al.

In this study, the authors present an investigation of the net flow of each bike-sharing station to quantify the station state. Then a bike-sharing demand prediction method based on autoregressive integrated moving average models is proposed. With a view of bicycles that need repair, an extended Vehicle Routing Problem with Simultaneous Picked and Delivery (VRPSPD) program is developed to model the bike-sharing static rebalancing problem. A hybrid Discrete Particle Swarm Optimization-Variable Neighborhood Search (DPSO-VNS) algorithm is employed to solve the proposed model.

“Impact of a New Metro Line: Analysis of Metro Passenger Flow and Travel Time Based on Smart Card Data” by X. Fu and Y. Gu

In this study, smart card data of metro system from Nanjing, China, are used to study the changes of metro passenger flow and travel time due to the operation of a new metro line. The impacts of the new metro line on passenger flow distribution, travel time in the metro network, and commuters' trips are analyzed.

“Improving Traffic State Prediction Model for Variable Speed Limit Control by Introducing Stochastic Supply and Demand” by Y. Bie et al.

This paper investigates how to eliminate this prediction error within a VSL environment. In this study, the traffic state prediction model is a second-order traffic flow model named METANET, while the VSL control is model predictive control (MPC) based, and the VSL decision is discrete

optimized choice. A simplified version of the switching mode stochastic cell transmission model (SCTM) is integrated with the METANET model to eliminate the prediction error. The performance of the proposed method is assessed using field data from a VSL pilot test in Edmonton, Canada, and is compared with the prediction results of the baseline METANET model during the road test.

“A Two-Layer Network Dynamic Congestion Pricing Based on Macroscopic Fundamental Diagram” by B. Wei and D. (J.) Sun

This paper proposes a bilevel programming toll model, incorporating MFD to solve the unbalanced flow distribution problem within the two-layer transportation networks. The upper-level model aims at minimizing the total travel time, while the lower level focuses on the MFD-based traffic assignment, which extends the link-based traffic assignment to network wide level. Genetic algorithm (GA) and the method of successive average were adopted for solving the proposed model, on which an online experimental platform was established using VISSIM, MATLAB, and Visual Studio software packages. The results of numerical studies demonstrate that the total travel time is decreased by imposing the dynamic toll, while the total travel time savings significantly outweigh the toll paid.

“Automatic Estimation Method for Intersection Saturation Flow Rate Based on Video Detector Data” by L. Wang et al.

In this research, an automatic estimation method is proposed for the Saturation Flow Rate (SFR) based on video detector data in order to overcome the limitation of the field measurement method. The actual vehicle headway is treated as time series and an auxiliary regression equation is built up wherein the parameters are estimated through the ordinary least square method. The SFR is eventually calculated using the average value of saturation headways. A case study is conducted to demonstrate the validity of the proposed model by using data from an intersection with three approaches in Qijing, China.

“Optimal Bus-Bridging Service under a Metro Station Disruption” by H. Yin et al.

This paper aims to solve the metro station disruption problem. The authors propose an integrated optimization model by providing additional bus-bridging services. The model includes an optimization part and a three-layer nested-logit passenger behavior model part for analyzing the dynamic passenger flow demand. A genetic algorithm and a simulation based method are integrated to solve the proposed model.

“A Crash Surrogate Metric Considering Traffic Flow Dynamics in a Motorway Corridor” by X. Wang and K. Liu

In this study, the authors propose a new concept of traffic state vulnerability to develop a simplified crash surrogate metric (SCSM) with a closed form in consideration of

traffic flow dynamics. A simulation model is developed using VISSIM to effectively generate dynamic vehicle behavior and a case study is conducted to compare the performance of the proposed metric with time to collision (TTC) and aggregated crash index (ACI). The result show that SCSM and ACI have similar performance which is superior to TTC, and SCSM is more applicable to practical engineering issues since it is less computational demanding.

“Optimal Design of Transportation Networks with Automated Vehicle Links and Congestion Pricing” by Y. Ye and H. Wang

This paper is focused on transportation network design problem (NDP) with congestion pricing involving both automated vehicles (AV) and conventional vehicles (CV) modes. This problem is formulated as a bilevel programming model. The upper-level optimizes the network design for AV links and congestion pricing while lower level describes a user equilibrium problem. The effectiveness of the proposed model and solution method are evaluated through numerical examples and some insights are obtained.

“Minimizing the Average Delay at Intersections via Presignals and Speed Control” by M. Ghanbarikarekani et al.

This paper aims to propose a model to improve presignals by reducing the vehicles' number of stops behind the presignals. By applying the method, vehicles would be able to adjust their speed based on traffic conditions as well as buses' speed and approach. Numerical analyses have been conducted to determine the conditions required for implementing this method.

“An Association Rule Based Method to Integrate Metro-Public Bicycle Smart Card Data for Trip Chain Analysis” by D. Zhao et al.

This research is focused on matching metro and public bicycle smart cards of the same commuters. To this end, a novel method with association rules to match the data derived from the two systems is proposed. The matching result helps to identify the spatial pattern of public bicycle demands, and the total journey and transit times of the metro-bicycle trip chain.

“On the Effects of Various Measures of Performance Selections on Simulation Model Calibration Performance” by C. Wang and C. Xu

In this study, the authors investigate how different performance measures affect the calibration of traffic simulation models at an intersection. In terms of reflecting traffic conditions and vehicle conflicts, they found calibrations on the basis of efficiency measures such as travel times and headway distribution performs better than those of safety measures and also a multimeasures based calibration performs better.

Collectively, these 20 papers illustrate the diverse range of the topics on transport system flow analysis and provide an insightful compilation recently being investigated in this

field. The papers included in this special issue are representatives of the current research challenges in advanced understanding, modelling, and algorithmic techniques for transport system flow analysis, which are capable of solving a variety of problems. It is expected that these papers can provide researchers with valuable inspiration and guidance to work on studies involving macroscopic fundamental diagram and its applications, disruptions to transport system flow, data-driven approaches for improving transport system planning, multimodal transport network flow, and so on.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

As guest editors, we would like to take this opportunity to thank all the authors for their contributions to this special issue and the reviewers for their expert review comments. We hope the readers will share our joy and find this special issue very helpful.

*Zhiyuan Liu
Lele Zhang
David Z.W. Wang*

Research Article

On the Effects of Various Measures of Performance Selections on Simulation Model Calibration Performance

Chen Wang ^{1,2} and Chengcheng Xu ¹

¹Jiangsu Key Laboratory of Urban ITS, Southeast University, Nanjing, 210096, China

²Intelligent Transportation Research Center, Southeast University, Nanjing, 210096, China

Correspondence should be addressed to Chengcheng Xu; xuchengcheng@seu.edu.cn

Received 27 May 2018; Revised 14 October 2018; Accepted 28 October 2018; Published 20 December 2018

Guest Editor: Lele Zhang

Copyright © 2018 Chen Wang and Chengcheng 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.

Objective. This paper examines the effects of various measures of performance (MOP) selections on simulation model calibration performance, in terms of reflecting actual traffic conditions and vehicle interactions. **Method.** Two intersections in Shanghai were selected for simulation model calibration, one for testing and another for validation. Three effective MOPs were utilized, including average travel time (i.e., time passing the intersection), average queue length, and vehicle headway distribution. The counts of three types of traffic conflicts (i.e., crossing, rear-end, and lane change) were used as safety MOPs. Those MOPs, as calibration objectives, were examined and compared. **Results.** The results of the testing site showed that different effective MOPs had their own advantages: average travel time appeared to be the best in reflecting lane change and rear-end conflicts while headway distribution performed the best consistency between simulated and actual crossing conflicts. Compared to the safety MOPs, average travel time and headway distribution still performed better, in terms of resulting in more similar simulated conflict metrics (e.g., TTC, PET) to actual ones. A multicriteria calibration strategy based on average travel time and headway distribution generally had better performances in reflecting actual traffic conditions and vehicle interactions than using any single effective or safety MOP. Similar results were found for the validation site. **Conclusion.** To simulate actual traffic conditions and vehicle interactions, multiple effective MOPs could be simultaneously considered for model calibration, instead of using safety MOPs.

1. Introduction

Microscopic traffic simulation has been widely utilized for traffic design, control, and management, since it is able to simulate actual traffic conditions so that unknown impacts of various traffic design and control scenarios can be properly evaluated. In order to develop a reliable microscopic simulation model, model calibration is considered as the most critical step that matches simulated traffic conditions with actual environment. To calibrate a simulation model, a measure of performance (MOP) usually needs to be selected as the calibration objective, such as average queue length [1], volume [2], speed [3], travel time (or time passing the intersection) [4], control delay [5], and headway distribution [6]. However, there seems to be no benchmark MOP for microscopic simulation model calibration. According to Park and Qi [4], a MOP was usually selected due to the ease of

data collection in the field. However, such method could also raise a question whether all MOP selections have comparable performances in reflecting actual traffic conditions. To our knowledge, there is limited research on the effects of MOP selections on simulating actual traffic conditions.

To note, most conventional MOPs are effective measures since simulation was initially conducted for traffic operational studies. Recently, microscopic simulation has been recognized for traffic safety analysis. To do this, it is important to reflect actual vehicle interactions. Previous research found that by incorporating a safety MOP (e.g., conflict counts) for calibration, simulation models appeared to better reflect actual vehicle interactions than calibrating effective MOPs [7–11]. However, such calibration strategy requires additional calibration efforts of safety-related data collection. For instance, in order to calibrate a simulation model based on traffic conflicts, conflicts need to be identified

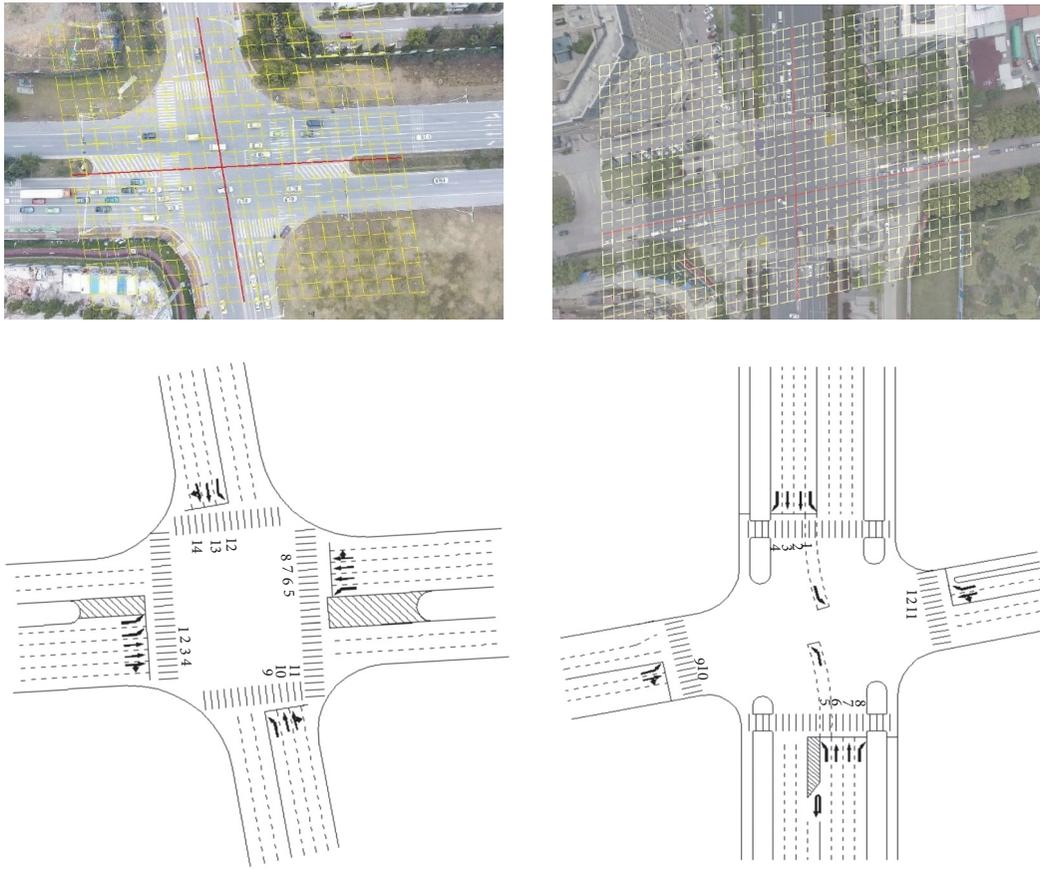


FIGURE 1: Aerial photos and sketches of the testing site (left) and the validation site (right).

and detailed conflict data need to be collected in the field. To note, collecting conflict data requires either a large amount of human observations or solid knowledge/skills of computer vision techniques. In practice, however, such requirement is very difficult to be fully fulfilled. Not to mention that video techniques often suffer from adverse weather, occlusions, and algorithm limitations. In general, safety-based calibration is difficult and requires a lot of efforts. If a calibrating strategy based on effective MOPs can have reasonable performances in simulating actual vehicle interactions, safety-related calibration efforts may not be required. Duong et al. [3] discussed this topic by proposing a calibration strategy based on two effective MOPs for freeway segments. The result showed the potential of using effective MOPs for simulating actual vehicle interactions. However, for urban intersections, there is no literature addressing such issue.

In light of these, it is necessary to find out how various MOPs (as calibration objectives) would affect the consistency between simulated and actual traffic conditions, as well as vehicle interactions. In this paper, we will examine the effects of various effective/safety MOP selections on simulation model calibration, in terms of reflecting actual traffic conditions and vehicle interactions at signalized intersections. The expectation of this research is to identify a promising calibration strategy that can reasonably reflect both actual traffic conditions and vehicle interactions.

2. Data Collection

Two signalized intersections in Fengxian District in Shanghai were selected in this study, one for testing and another for validation. At the initial stage of this study, video clips were collected by portable cameras, set up on roadside high-rise buildings [11]. This set-up provided enough view coverage to collect most data needed, including traffic conflicts, volume, and headway. However, to examine the effects of various MOPs on simulating actual traffic conditions, average queue length is an important effective MOP widely used in previous literature. Thus, to get enough viewing height to cover whole intersection area (shown in Figure 1), a drone was used to collect video clips during afternoon peak hours (4:30-6:30) on two different weekdays.

Conflicts were identified based on traditional TCT methods, by four trained observers. Those observers were early examined their intrareliability and interreliability, by watching other video clips [11]. A clear sign of evasive actions (e.g., braking, swerving) was used as the sign of conflicts. A grid system was placed in the video to match global coordinates with local coordinates, by which vehicle positions and distance can be manually estimated. Video clips were viewed at 25 frames per second to estimate moving velocity of vehicles, using VideoStudio. Furthermore, conflict metrics were calculated by observers, including Time-to-Collision (TTC) and Postencroachment Time (PET).

It should be noted that conflicts could be identified by automated video techniques, including unsupervised or semisupervised learning algorithms [12]. Unsupervised learning algorithms could quickly differentiate conflicts without manually labeling data. However, it often suffers from a low detection accuracy issue [12]. On the other hand, supervised or semisupervised learning algorithms could better identify conflicts but require many efforts to collect training data. In general, there is an obvious trade-off between efficiency and accuracy, regarding video conflict identification and data collection. Moreover, those algorithms still inherently depend on human judgment on conflict identification for either data labeling or validation. Thus, from any perspective, conflict identification could be a difficult task for middle-to large-scale microscopic traffic simulation calibration. The derivation of vehicle speed and conflict metrics may be improved by extracting high-resolution vehicle trajectories based on automated video techniques [13, 14]. However, the reliability of such method still suffers from adverse weather, varying light conditions, and occlusion.

Thus, in our study, around over ten hours of manual calculation efforts were conducted to obtain those detailed conflict data, for each intersection. This data collection process was believed to adequately capture actual traffic conditions and vehicle interactions in the real world. The labor-assist conflict identification approach was believed to be a more straightforward and reliable way to identify conflicts in this study. Admittedly, when considering larger number of intersections, it will become labor-intensive to identify conflicts by human efforts. However, it also highlights the necessity of finding other effective MOPs comparable to safety MOPs (e.g., conflicts), for microscopic traffic simulation calibration.

Figure 1 presents the aerial photos of the two intersections, as well as their sketches with numbered lanes. All effective MOPs were calculated by each lane (i.e. lane-based) and aggregated into one hour. Thus, for the testing site, there were 14 (lanes) * 4 (hours) = 56 samples. For the validation site, there were 12 (lanes) * 4 (hours) = 48 samples. Conflicts are not considered as lane-specific, because many are difficult to be assigned into a single lane (e.g., crossing and lane change conflicts). Instead, they were analyzed by total counts, PET, and TTC. Table 2 presents the summary statistics of traffic and conflict data collected in this study. Detailed lane-specific traffic data were included in the Appendices A and B.

3. Simulation Model Development

In this study, a commercial microscopic simulation package VISSIM was used to develop simulation models for the two intersections. VISSIM has been extensively used in simulation studies [9–11, 15–17], due to its high flexibility in simulating actual traffic conditions, especially microscopic driver behaviors [18, 19]. Geometric data were collected based on aerial photographs captured by drone and field observations. Technical drawings were also provided by the Fengxian Highway Administration, with detailed roadway geometrics including the number and width of lanes, length of storage bays and tapers, and details of lane utilization.

Traffic control details were derived from the Fengxian Traffic Police Department and field observation, including posted speed limits, signal timing plans, and movement permissions. Traffic flow data were mostly collected based on videos, including traffic counts and vehicle composition. Those data were used as the input for developing a base simulation model for each intersection.

As for simulation calibration, many MOPs have been used in previous studies [7], including average delay, average travel time, number of stops, headway distribution, and average queue length. In this study, average travel time (time passing an intersection), headway distribution, and average queue length were utilized as effective MOP. Each was examined for their effects on reflecting actual traffic conditions. In previous literature, conflict counts were often used as a safety MOP to reflect actual vehicle interactions. Thus, in this study, conflict counts were used for model calibration, while TTC and PET were used to verify calibrated simulation models.

4. Simulation Model Calibration for Testing Site

The testing site was used to examine the effects of various MOP selections on reflecting actual traffic condition and vehicle interactions. Effective MOPs include average travel time, average queue length, and headway distribution. Conflicts were classified into three types: crossing, rear-end, and lane change. The counts of each type were used as a safety MOP, resulting in three safety MOPs. Surrogate Safety Assessment Model (SSAM) was used to extract three types of simulated conflicts, with the suggested angle thresholds (30° and 85°) [20]. As for PET/TTC thresholds, 3 seconds was used because most actual observations fell within this range (as shown in Table 1).

First of all, a set of ANOVA tests were applied to identify sensitivity parameters to each effective and safety MOP. Since the two intersections are in urban area, parameters in Wiedemann 74 model were selected for model calibration. The list and the acceptable ranges of driver behavior parameters in VISSIM were determined initially based on previous literatures and a Latin Hypercube Sampling- (LHS-) based calibration approach was early adopted to ensure that the full range of parameter combinations was sampled [21]. Based on ANOVA tests, a set of parameters were found as sensitive. Table 2 presents the results.

From Table 2, some parameters were found to be sensitive to both effective and safety MOPs. For example, minimum gap time was found to be sensitive to both headway distribution and crossing conflict counts. This result was consistent with previous literature [7]. Average standstill distance, additive, and multiplicative part of desired safety distance affected average travel time, average queue length, and rear-end conflict counts. Those parameters could reflect interactions between leading and following vehicles. Safety distance reduction factor and min headway were found as sensitive to average travel time and lane change conflict counts. This could be reasonable since those parameters mainly control lane change behaviors. Another finding is that different effective MOPs have different sensitive parameters.

TABLE 1: Summary statistics of collected traffic and conflict data.

Lane-based hourly effective MOP		Testing Site				Validation Site			
		mean	s.d.	min	max	mean	s.d.	min	max
Volume		247	16.2	82	497	222	15.8	61	435
Queue length		67.2	3.4	17	107	63.9	3.0	16	102
Headway		2.2	0.1	1.1	3.8	2.4	0.12	1.1	4
Travel time		5.5	0.24	2.7	10.9	5.0	0.22	2.6	8.0
Safety MOP		Testing Site				Validation Site			
	Counts	72				62			
crossing conflicts	TTC	1.5	0.61	0.5	2.9	1.5	0.58	0.6	3.3
	PET	1.30	0.63	0.3	3.0	1.33	0.66	0.3	3.2
	Counts	165				141			
rear-end conflicts	TTC	1.5	0.53	0.4	3.2	1.57	0.62	0.5	3.2
	PET	1.23	0.57	0.3	3.0	1.37	0.61	0.3	3.0
	Counts	56				45			
lane change conflicts	TTC	1.41	0.63	0.4	3.1	1.54	0.68	0.4	3.5
	PET	1.31	0.62	0.4	3.3	1.38	0.72	0.3	3.2

TABLE 2: ANOVA tests on sensitive driving behavior parameters to MOPs.

	average travel time	Average queue length	Headway distribution	crossing conflict counts	rear-end conflict counts	lane-change conflict counts
Number of observed preceding vehicles	0.001	0.323	0.442	0.342	0.532	0.326
average standstill distance	0.001	0.001	0.217	0.634	0.001	0.255
additive part of desired safety distance	0.323	0.003	0.557	0.627	0.001	0.137
multiple part of desired safety distance	0.455	0.001	0.105	0.246	0.001	0.631
waiting time before diffusion	0.727	0.001	0.075	0.442	0.552	0.547
Min headway (front/rear)	0.003	0.562	0.224	0.515	0.872	0.001
Acceptable deceleration of trailing vehicle	0.532	0.722	0.001	0.332	0.239	0.115
safety distance reduction factor	0.005	0.156	0.137	0.634	0.772	0.001
minimum headway	0.001	0.346	0.379	0.184	0.592	0.219
Minimum gap time	0.113	0.723	0.001	0.001	0.417	0.321

Moreover, sensitive parameters to safety MOPs are also varying.

Then, Genetic Algorithm (GA) was applied to calibrate the simulation model based on each MOP selection, respectively. GA is an adaptive heuristic search algorithm, which was commonly used in simulation calibration to find optimized driving behavior parameters that could reflect actual traffic conditions to the largest extent [7]. GA includes four main steps: (1) Population generation; (2) Selection; (3) Crossover; and (4) Mutation. Normally, initial population needs to be randomly generated. In this case, a random combination of sensitive parameters was considered as an individual of GA. Ten individuals were included in the initial population, with random population generation. A selection operator is used for giving preference to better individuals, allowing them to pass on their genes to the next generation. Normally, a fitness function is created to determine the goodness of individuals. Regarding simulation

calibration, various indicators can be used, such as mean average percentage error (MAPE), root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and the Pearson correlation coefficient. To note, in our study, MAPE was used for developing fitness functions, while all other listed indicators were calculated for verifying calibration performances. The equation of MAPE, RMSE, MSE, MAE, and correlation coefficients can be written as

$$\text{MAPE} = \frac{1}{N} \sum_{k=1}^N \frac{|C_k - A_k|}{A_k} \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^N (C_k - A_k)^2} \quad (2)$$

$$\text{MSE} = \frac{1}{N} \sum_{k=1}^N (C_k - A_k)^2 \quad (3)$$

TABLE 3: Calibrated driving behavior parameters for testing site.

	average travel time	Average queue length	headway distribution	crossing conflict counts	rear-end conflict counts	Lane change conflict counts	Headway& travel time
Number of observed preceding vehicles	4*	2	2	2	2	2	3
average standstill distance	2*	1.75*	2	2	1.75*	2	1.80*
additive part of desired safety distance	3	2.85*	3	3	2.65*	3	3
multiple part of desired safety distance	3	2.15*	3	3	2.35*	3	3
waiting time before diffusion	60	30*	60	60	60	60	60
Min headway (front/rear)	1.42*	1.64	1.64	1.64	1.64	1.55*	1.45*
Acceptable deceleration of trailing vehicle	-0.5	-0.5	-0.8*	-0.5	-0.5	-0.5	-0.8
safety distance reduction factor	0.73*	0.6	0.6	0.6	0.6	0.88*	0.75*
minimum headway (priority rules)	3.8*	5	5	5	5	5	3.9*
Minimum gap (priority rules)	3	3	2.45*	2.70*	3	3	2.50*

*indicates calibrated values.

$$MAE = \frac{1}{N} \sum_{k=1}^N |C_k - A_k| \quad (4)$$

$$Correlation = \frac{\sum_{k=1}^N (C_k - \bar{C})(A_k - \bar{A})}{\sqrt{\sum_{k=1}^N (C_k - \bar{C})^2} \sqrt{\sum_{k=1}^N (A_k - \bar{A})^2}} \quad (5)$$

where N is total sample points; C_k is the calibrated MOP for the k_{th} sample point; A_k is the actual MOP for the k_{th} sample point; \bar{C} , \bar{A} are the average value of the calibrated/actual MOP of all sample points, respectively.

After that, a crossover operator needs to be applied to create one new “offspring” individual based on two “parent” individuals. That is, two sets of parameter values were used to create a new parameter set, which had values from either of them. The last step is mutation. A portion of new individuals (i.e., parameter combinations) will be randomly selected and modified, to maintain diversity within the population and inhibit premature convergence. This process was repeated until the convergence of the calibration. In this study, population size, crossover rate, and mutation rate were set as 10, 0.6, and 0.3, respectively. All calibration processes converged within 80 generations.

Table 3 shows the optimal parameter values after GA convergence, for each MOP selection. Note that a multicriteria calibration result based on headway and average travel time was also reported in Table 2. Based on the results of each single effective MOP, a multicriteria calibration strategy based on effective MOPs was further applied. Each MOP was set as a weight and the calibration objective was set as

$$\min X = w_1 z_1(X) + w_2 z_2(X) \quad (6)$$

where w_i is the weight for the i_{th} MOP; z_i is the performance indicator of the i_{th} MOP (e.g., MAPE of queue length); X is the vector of parameters to be calibrated.

Multicriteria optimization problem has been studied for years [22, 23]. In this study, equal weights were used for GA calibration. Among multicriteria scenarios (i.e., travel time and queue length, travel time & headway, headway and queue length), the scenario based on headway and travel time showed the best overall calibration performance.

Figure 2 illustrates the whole calibration procedure applied in the study.

To examine the effects of different MOPs on calibration performances, both effective and safety perspectives were considered. Effective measures for performance verification are the same as those for calibration. For each conflict type, three safety-based measures were selected to verify calibrating performance, including conflict counts, average TTC, and average PET. In most literature, only conflict counts were often used for model verification. However, to accurately reflect actual vehicle interactions, detailed conflict metrics were also believed as important. Table 4 presents the verification of calibration performances, in terms of MAPE. Note that, for safety-based calibration, only the best performances of each conflict type were presented.

For crossing conflicts, the safety-based calibration strategy resulted in the lowest MAPE (29%) for total conflict counts. This is reasonable since conflict counts are the direct calibrating objective for safety-based calibration. Among effective MOPs, headway distribution appeared to be a better calibrating objective than average travel time and average queue length, in terms of presenting 34% MAPE for crossing conflict counts. Moreover, it has 18% MAPE for average TTC and 23% MAPE for average PET, which is better than safety

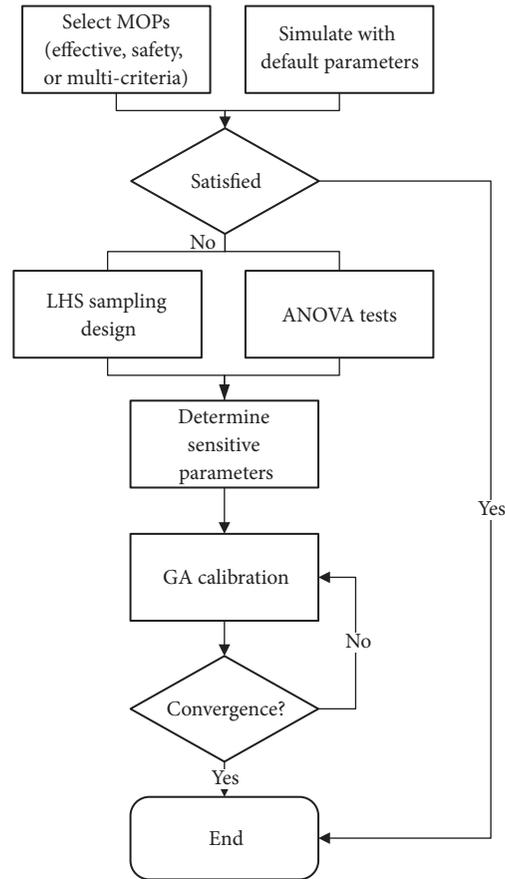


FIGURE 2: A general simulation calibration procedure applied in the study.

MOPs. For rear-end conflicts, calibration based on the safety MOP (i.e., conflict counts) resulted in the lowest MAPE (21.4%) for conflict counts, followed with average travel time (24.2%). However, as for average TTC and PET, average travel time appeared to be the best MOP among all, with the lowest MAPEs. For lane change conflicts, average travel time was shown as the best MOP for calibration, with the lowest MAPE for conflict counts, average TTC, and average PET. On the other hand, calibrating safety MOPs appeared to be unable to reflect actual vehicle interactions, in terms of presenting relatively high MAPEs for detailed safety-based verification measures (i.e., average TTC and PET).

It is reasonable that safety-based calibration could generally result in lower MAPEs for conflict counts. Previous literature also reported similar results and claimed the advantage of such strategies over conventional effective-based calibration strategies. However, in this study, safety-based strategies were found with lower MAPEs for average TTC and PET.

Headway distribution appeared to be the best calibrating objective for crossing conflicts, in terms of presenting lowest MAPEs for average TTC and PET. However, it did not perform very well in reflecting the other two types of conflicts. Average travel time better reflected actual vehicle interactions in terms of rear-end conflicts and lane change

conflicts. However, it provided relatively high MAPEs for crossing conflicts. Average queue length generally resulted in high MAPEs for all three conflict types. Thus, headway distribution and average travel time appeared to be complementary in matching simulated vehicle interactions with actual vehicle interactions. More importantly, they are comparable or even better than safety MOPs used for calibration in this study. Thus, it is reasonable to expect a multicriteria calibration based on two effective MOPs can be better than a single effective MOP and safety MOP. This was proven by the multicriteria calibration results (i.e., headway distribution and average travel time), which provided reasonable MAPEs for all verification measures. Even some MAPEs are slightly higher than calibration based on single effective/safety MOPs, the results can still be considered as acceptable due to the randomness of simulation nature as well as the potential trade-off between two effective MOPs.

5. Calibration Results for Validation Site

To examine the transferability of the effects of MOP selections on simulation calibration performances, another intersection was examined based on the same calibration procedure. First, sensitivity parameters were found as comparable to the testing site. Then, calibration efforts based on single effective

TABLE 4: The verification of model calibration performance for testing site (in MAPE).

MOP (verification)		travel time	queue length	MOP (calibration)		
				Headway distribution	Safety MOP	Headway & travel time
crossing conflict	Counts	44%	56%	34%	29%	31%
	Average TTC	42%	58%	18%	33%	27%
	Average PET	34%	41%	23%	37%	21%
	Counts	24.2%	30%	34%	21.4%	23%
	Average TTC	33%	41%	39%	42%	15%
	Average PET	27%	45%	36%	48%	22%
lane change conflict	Counts	34%	58%	54%	47%	38%
	Average TTC	32%	60%	58%	65%	26%
	Average PET	37%	65%	62%	59%	34%
Effective Measures	Average Travel Time	17%	25%	24%	38%	17%
	Average Queue Length	29%	19%	33%	48%	24%
	Headway Distribution	20%	22%	11%	41%	16%

TABLE 5: Calibrated driving behavior parameters for validation site.

	average travel time	Average queue length	headway distribution	crossing conflict count	rear-end conflict count	Lane change conflict count	Headway& travel time
Number of observed preceding vehicles	4*	2	2	2	2	2	3
average standstill distance	2.2*	1.80*	2	2	1.80*	2	1.9*
additive part of desired safety distance	3	2.9*	3	3	2.65*	3	3
multiple part of desired safety distance	3	2.15*	3	3	2.4*	3	3
waiting time before diffusion	60	40*	60	60	60	60	60
Min headway (front/rear)	1.5*	1.64	1.64	1.64	1.64	1.6*	1.55*
Acceptable deceleration of trailing vehicle	-0.5	-0.5	-0.8*	-0.5	-0.5	-0.5	-0.8
safety distance reduction factor	0.7*	0.6	0.6	0.6	0.6	0.9*	0.80*
minimum headway (priority rules)	3.9*	5	5	5	5	5	4.0*
Minimum gap (priority rules)	3	3	2.50*	2.85*	3	3	2.55*

*indicates calibrated values.

MOPs and safety MOPs were conducted. Calibration details are shown in Table 5. Similarly, safety-based calibration was unable to capture detailed vehicle interactions, in terms of average TTC and PET. As for effective MOPs, average travel time appeared to be the best calibrating objective for lane change conflicts and headway distribution outperformed other MOPs in terms of having lowest MAPE for crossing conflicts. For rear-end conflicts, average queue length and average travel time had comparable MAPEs. Average queue length had slightly better MAPEs for average PET and TTC.

Due to those effects, a multicriteria calibration effort was conducted based on average travel time and headway distribution. The results were consistent with the testing site. Again, the multicriteria calibration appeared to find a balance between two single effective MOPs by providing acceptable performances in reflecting all types of actual vehicle interactions (i.e., crossing, rear-end, and lane change). It also outperformed the safety-based calibration strategies. Calibration performances were reported in Table 6.

To further examine calibration performances, a lane-based test was also applied. Figure 3 shows the simulated observations (based on the multicriteria calibration) versus actual observations, for each lane with a one-hour aggregation. All performance indicators were reported, including MAPE, correlation, RMSE, MSE, and MAE. Generally, the multicriteria calibration reasonably reflected lane-based MOPs, for both sites.

Additionally, the TTC/PET distributions of actual conflicts, safety-based calibration results, and the multicriteria calibration were compared, for the two sites. Details are shown in Appendices C and D. It can be found that the multicriteria calibration resulted in simulated TTC and PET distribution, which are more similar to actual TTC/PET

distribution than safety-based calibrations. Chi-square tests were also conducted. It was found that there was no significant difference between simulated and actual TTC/PET in most cases, for the multicriteria calibration strategy.

6. Conclusion

This paper examines the effects of various MOP selections on simulation model calibration performances, in terms of reflecting actual traffic flow conditions and vehicle interactions at signalized intersections. Two intersections in Fengxian, Shanghai, were selected in the study: one site was used for testing those effects and another site was used for validation. Average travel time (i.e., time passing the intersection), average queue length, and headway distribution were used as effective MOPs for model calibration, which were also commonly used in previous literature. Actual vehicle interactions were represented by traffic conflict data collected in the field. Conflicts were carefully determined by four reliable trained observers, while detailed conflict metrics (e.g., TTC, PET) were manually calculated. The counts of three conflict types (i.e., crossing, rear-end, and lane change) were selected as safety MOPs for simulation calibration. A commercial simulation package, VISSIM, was utilized to develop base simulation models for the two intersections, based on fundamental geometric, traffic, and control data.

For the testing site, driving behavior parameters in VISSIM were examined for their sensitivity to the effective MOPs, as well as safety MOPs. An LHS-based calibration approach was early adopted to ensure that the full range of parameter combinations was sampled. In doing so, ANOVA tests can be properly applied to examine sensitivity of each parameter. Genetic algorithm (GA) was utilized to calibrate

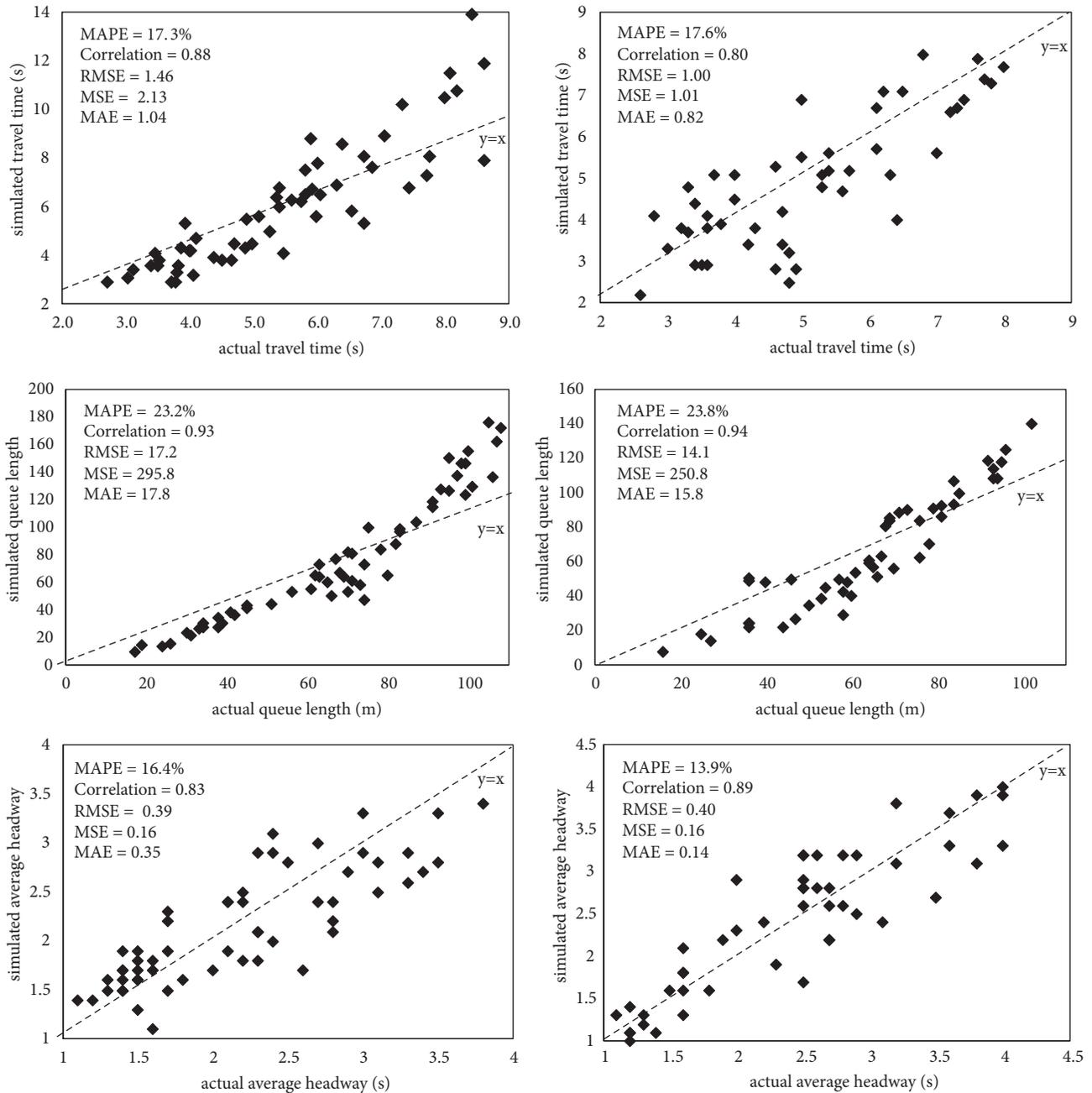


FIGURE 3: Simulated lane-based observations versus actual observations at each time interval at testing site (left) and validation site (right).

the microscopic simulation model and comparisons among different MOPs were conducted, in terms of examining the consistency between simulated outputs and actual data. The results showed that average travel time outperformed other effective MOPs, in terms of presenting the lowest MAPEs of average TTC and PET. Headway distribution was found to better reflect crossing vehicle interactions, in terms of producing lower MAPEs of average TTC and PET for crossing conflicts. On the other hand, only using conflict counts for model calibration did not result in higher consistency between simulated and actual vehicle interactions, especially for detailed conflict metrics (i.e., average TTC and

PET). A multicriteria calibration strategy based on headway and average travel time showed a considerable consistency between simulated outputs and actual data. Although it did not outperform single effective MOPs or safety MOPs in every aspect, it was considered as a promising selection to simulate actual traffic conditions and vehicle interactions. To further examine the transferability of those effects, another intersection was applied with the same calibration process. Similar results were also found.

Based on the results, some conclusions can be drawn as follows: (1) different effective MOPs may have different sensitivity parameters for model calibration; (2) different

TABLE 6: The verification of model calibration performance for validation site (in MAPE).

MOP (verification)		travel time	queue length	MOP (calibration)		
				Headway distribution	Safety MOP	Headway & travel time
Safety Measures	Counts	46%	48%	34%	26%	35%
	crossing conflict	Average TTC 38%	42%	23%	35%	29%
		Average PET 35%	40%	27%	38%	32%
	rear-end conflict	Counts 25%	33%	39%	20%	27%
		Average TTC 30%	35%	31%	39%	25%
		Average PET 27%	38%	26%	43%	26%
Effective Measures	lane change conflict	Counts 34%	43%	37%	35%	34%
		Average TTC 30%	47%	39%	51%	33%
		Average PET 31%	52%	42%	55%	31%
Effective Measures	Average Travel Time	19%	24%	25%	32%	18%
	Average Queue Length	26%	20%	31%	40%	24%
	Headway Distribution	17%	25%	13%	44%	14%

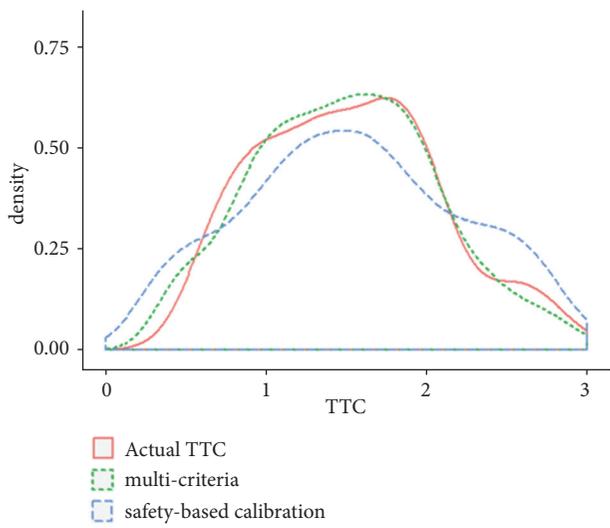
TABLE 7: Lane-based measures of performance at the testing site.

Lane No.	Hour 1			Hour 2			Hour 3			Hour 4						
	TT	QL	TH	VOL	TT	QL	TH	VOL	TT	QL	TH	VOL	TT	QL	TH	VOL
1	3.8	45	3.5	114	6.0	74	2.4	223	4.9	83	2.4	194	7.4	95	2.3	238
2	3.1	61	2.8	226	10.6	99	1.6	485	4.9	98	1.7	380	8.0	107	1.8	490
3	4.1	58	2.7	212	8.6	95	1.3	497	4.7	91	1.5	457	10.9	101	1.5	468
4	2.7	44	3.3	113	5.4	83	2.3	134	4.6	71	2.3	127	5.9	73	2.2	148
5	3.4	70	3.5	196	5.4	101	2.2	289	3.9	68	2.8	232	7.1	78	2.0	314
6	4.0	65	2.8	216	8.4	95	1.5	491	4.1	93	2.5	258	7.7	97	1.1	470
7	3.5	67	2.9	219	8.1	106	1.4	415	4.4	91	2.1	272	7.8	104	1.2	406
8	3.1	41	3.4	126	7.3	75	1.7	231	4	42	2.7	149	6.9	70	1.7	208
9	3.5	19	3.1	82	6.7	62	1.5	230	4.5	33	2.4	105	6.5	64	1.6	261
10	3.8	34	3.3	111	6.3	70	1.4	273	5.3	39	2.1	180	6.7	87	1.5	358
11	3.5	30	3.0	113	6.1	62	1.4	252	5.0	31	2.2	142	6.4	82	1.4	286
12	3	17	3.8	85	5.8	51	1.5	153	5.1	26	2.1	107	5.6	98	1.7	237
13	3.9	38	3.1	149	5.8	70	1.4	292	5.4	24	2.6	205	6.0	67	1.4	398
14	3.8	34	3.0	126	5.8	66	1.5	224	5.5	38	2.3	147	5.9	81	1.6	330

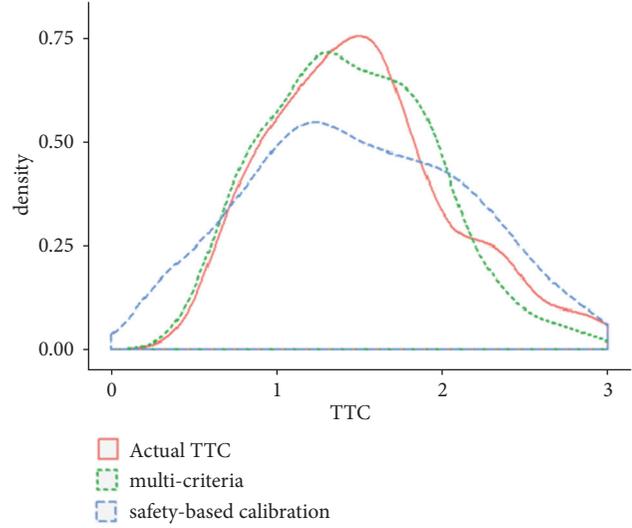
TABLE 8: Lane-based measures of performance at the validation site.

Lane No.	Hour 1				Hour 2				Hour 3				Hour 4							
	TT	QL	TH	VOL	TT	QL	TH	VOL	TT	QL	TH	VOL	TT	QL	TH	VOL	TT	QL	TH	VOL
1	3.4	102	4	112	5.4	79	2.3	205	4.7	65	2.7	170	6.8	47	2.5	203				
2	3.4	96	3.1	219	8	78	1.8	412	4.8	64	2	320	7.4	46	1.6	435				
3	3.8	95	3.2	209	7.7	76	1.6	428	4.6	64	1.6	377	7.6	44	1.6	422				
4	2.6	94	3.8	86	5	76	2.5	124	4.3	61	2.8	112	5.6	40	2.5	131				
5	3.2	93	3.6	160	4.8	73	2.4	257	3.7	60	2.9	204	6.4	36	2.2	273				
6	3.5	93	3.1	183	7.7	71	1.2	401	4	59	2.6	217	7	36	1.3	432				
7	3.3	92	3.2	204	7.3	70	1.6	369	4	58	2.5	239	7.2	36	1.6	371				
8	2.8	85	3.8	99	6.5	69	1.9	191	3.6	58	2.8	126	6.3	36	2	180				
9	3.3	84	3.6	68	6.1	69	1.2	197	4.2	57	2.5	97	6.2	36	1.2	226				
10	3.6	84	3.5	107	5.7	68	1.6	231	4.6	54	2.6	141	6.1	27	1.3	316				
11	3	81	4	61	5.3	67	1.6	142	4.7	53	2.5	96	5	25	1.2	209				
12	3.6	81	3.6	131	5.3	66	1.8	251	4.9	50	2.7	169	5.4	16	1.1	362				

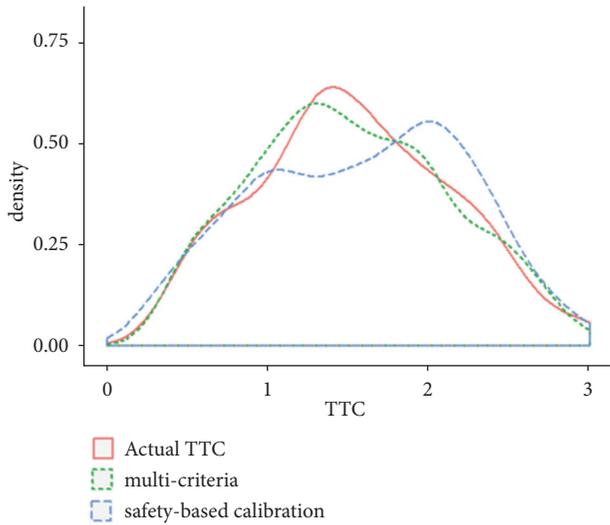
TT: travel time (s), QL: average queue length (m), TH: time headway (s), and VOL: traffic volume (veh/h)



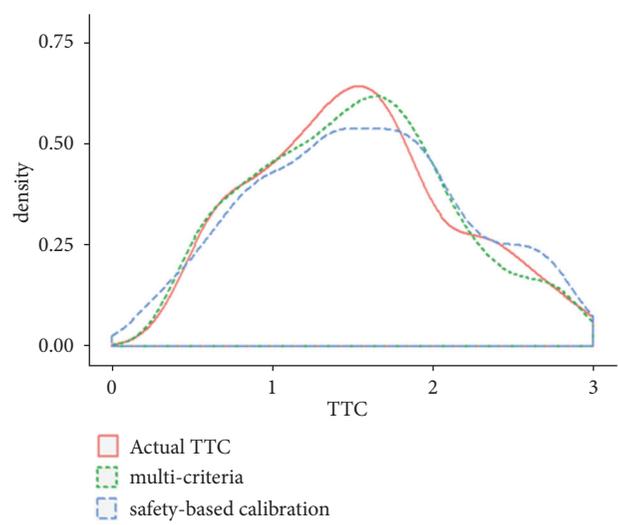
(a) Crossing Conflicts (Intersection 1)



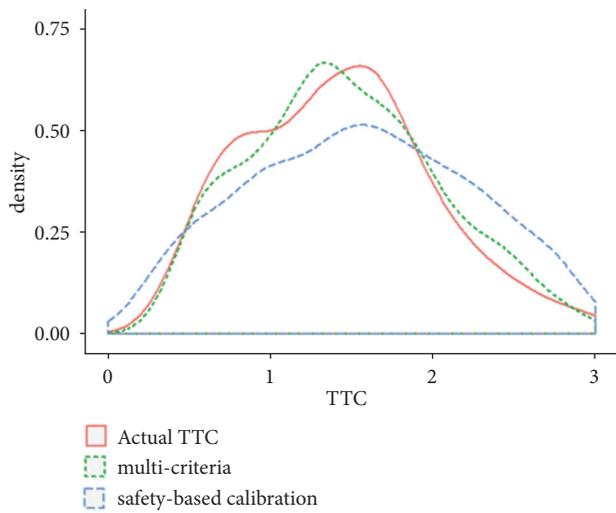
(d) Crossing Conflicts (Intersection 2)



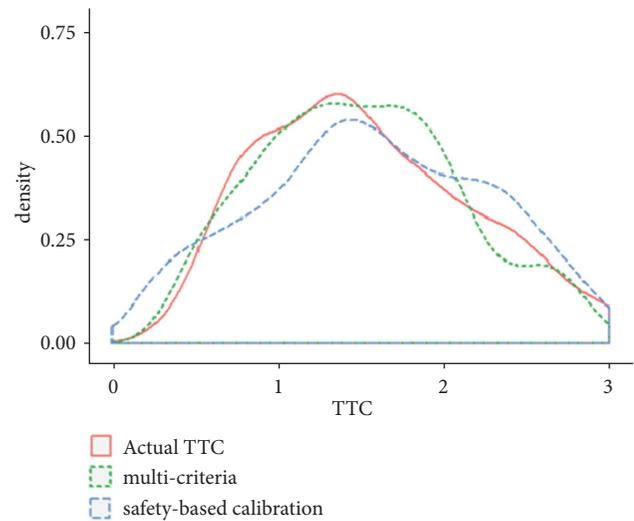
(b) Rear-end Conflicts (Intersection 1)



(e) Rear-end Conflicts (intersection 2)



(c) Lane Change Conflicts (Intersection 1)



(f) Lane Change Conflicts (Intersection 2)

FIGURE 4: TTC distribution of multicriteria calibration, safety-based calibration, and the field.

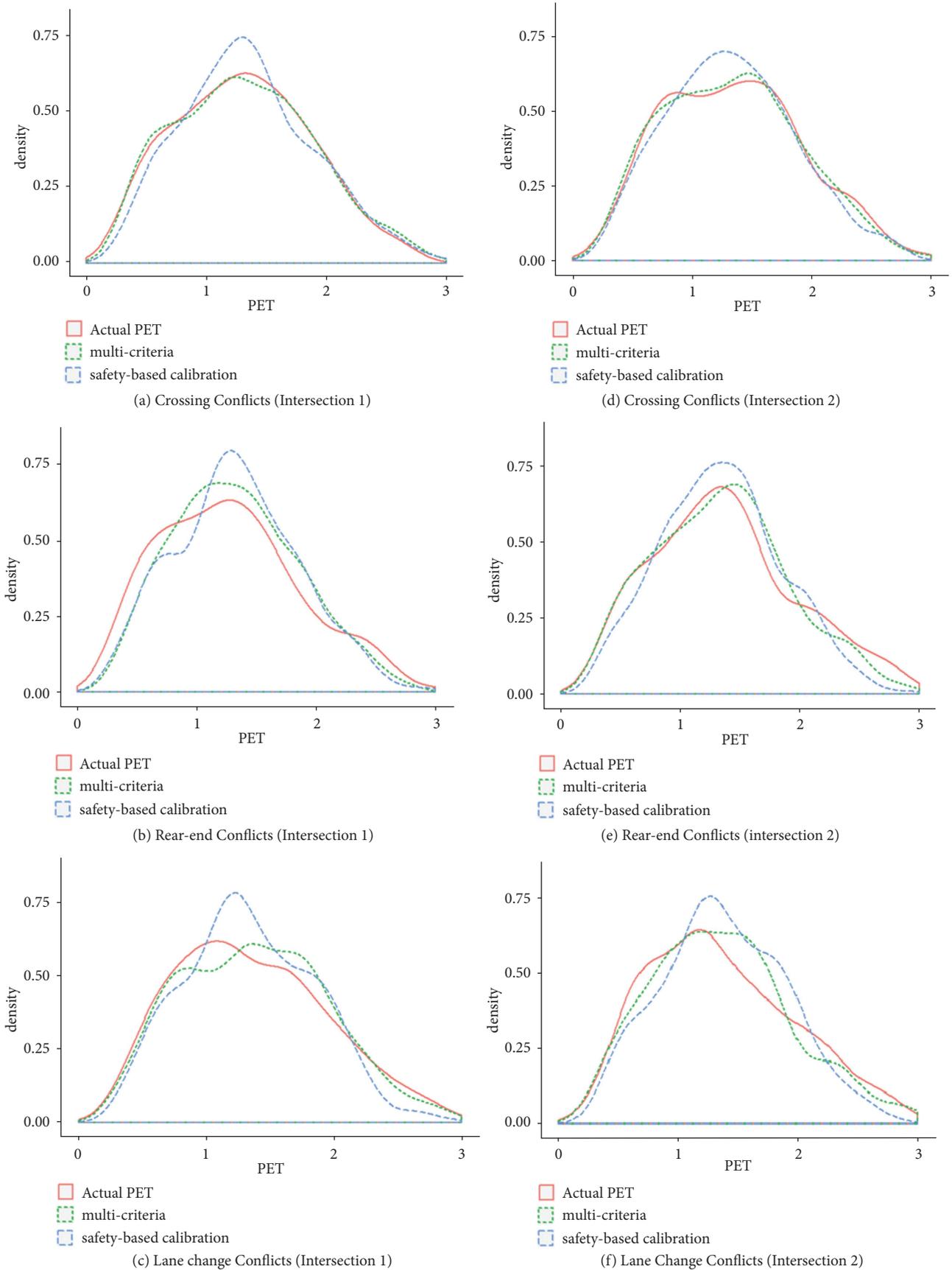


FIGURE 5: PET distribution of multicriteria calibration, safety-based calibration, and the field.

effective MOPs have their own advantages of reflecting actual vehicle interactions; (3) conflict counts may not be the optimal safety MOP for simulation calibration, in terms of reflecting actual vehicle interactions; (4) compared to conflict counts, using multiple MOPs resulted in better calibration performances in reflecting vehicle interactions, especially for detailed conflict metrics (i.e., TTC and PET). Thus, a multicriteria calibration strategy based on effective MOPs could be considered to simulate both actual traffic conditions and vehicle interactions.

Admittedly, this research also has some issues that need to be further addressed. First, only two intersections were considered in this study. It may not be enough to determine the transferability of the results. However, this was partly due to the difficulty of collecting safety MOPs, which also highlighted the necessity of calibrating proper effective MOPs as surrogates. Future validation studies could incorporate more intersections and more data collection efforts. Second, more complicated calibration strategies could be examined in the future. Currently, equal weights were used for calibration. However, it is interesting to identify the effect of unequal weights on calibration. Also, other safety-based calibration methods could also be compared, such as the two-stage calibration strategy [7]. Third, more effective MOPs could be examined, such as average number of stops. In addition, traffic facilities other than signalized intersections could be examined to validate the transferability of the effects of various MOPs as calibration objectives. We recommended that future research could be focused on these topics.

Appendix

A.

See Table 7.

B.

See Table 8.

C.

See Figure 4.

D.

See Figure 5.

Data Availability

The data used to support the findings of this study are available from the authors upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant nos. 51508093, 51608114, and 51408322) and the Fundamental Research Funds for the Central Universities #2242018K40008.

References

- [1] V. Gallelli, T. Iuele, R. Vaiana, and A. Vitale, "Investigating the transferability of calibrated microsimulation parameters for operational performance analysis in roundabouts," *Journal of Advanced Transportation*, vol. 2017, Article ID 3078063, 10 pages, 2017.
- [2] J. Hourdakakis, P. Michalopoulos, and J. Kottommannil, "A practical procedure for calibrating microscopic traffic simulation models," in *Proceedings of the TRB 82nd Annual Meeting Compendium of Papers*, vol. 1852, January 2003.
- [3] D. Duong, F. Saccomanno, and B. Hellinga, "Calibration of microscopic traffic model for simulating safety performance," in *Proceedings of the TRB 89th Annual Meeting Compendium of Papers*, January 2010.
- [4] B. Park and H. Qi, "Development and evaluation of a procedure for the calibration of simulation models," *Transportation Research Record*, vol. 1934, pp. 208–217, 2005.
- [5] Y. Zhang, J. Sun, and A. Kondyli, "An empirical framework for intersection optimization based on uniform design," *Journal of Advanced Transportation*, vol. 2017, Article ID 7396250, 10 pages, 2017.
- [6] C. Li, M. Karimi, and C. Alecsandru, "Microscopic simulation-based high occupancy vehicle lane safety and operation assessment: a case study," *Journal of Advanced Transportation*, vol. 2018, Article ID 5262514, 12 pages, 2018.
- [7] F. Huang, P. Liu, H. Yu, and W. Wang, "Identifying if VISSIM simulation model and SSAM provide reasonable estimates for field measured traffic conflicts at signalized intersections," *Accident Analysis & Prevention*, vol. 50, pp. 1014–1024, 2013.
- [8] R. Fan, H. Yu, P. Liu, and W. Wang, "Using VISSIM simulation model and surrogate safety assessment model for estimating field measured traffic conflicts at freeway merge areas," *IET Intelligent Transport Systems*, vol. 7, no. 1, pp. 68–77, 2013.
- [9] E. Mohamed and T. Sayed, "Transferability of calibrated microsimulation model parameters for safety assessment using simulated conflicts," *Accident Analysis and Prevention*, vol. 84, pp. 41–53, 2015.
- [10] E. Mohamed and T. Sayed, "A comparison between PARAMICS and VISSIM in estimating automated field-measured traffic conflicts at signalized intersections," *Journal of Advanced Transportation*, vol. 50, no. 5, pp. 897–917, 2016.
- [11] C. Wang, C. Xu, J. Xia, Z. Qian, and L. Lu, "A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 281–291, 2018.
- [12] S. Nicolas and T. Sayed, "A probabilistic framework for the automated analysis of the exposure to road collision," in *Proceedings of the TRB 87th Annual Meeting Compendium of Papers*, January 2008.
- [13] R. Danescu, F. Oniga, S. Nedevschi, and M. Meinecke, "Tracking multiple objects using particle filters and digital elevation maps," in *Proceedings of the 2009 IEEE Intelligent Vehicles Symposium (IV)*, pp. 88–93, Xi'an, China, June 2009.

- [14] Z. Cai, Z. Gu, Z. L. Yu, H. Liu, and K. Zhang, "A real-time visual object tracking system based on Kalman filter and MB-LBP feature matching," *Multimedia Tools and Applications*, vol. 75, no. 4, pp. 2393–2409, 2016.
- [15] F. Cunto and F. F. Saccomanno, "Calibration and validation of simulated vehicle safety performance at signalized intersections," *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1171–1179, 2008.
- [16] C. Wang and N. Stamatidis, "Evaluation of a simulation-based surrogate safety metric," *Accident Analysis & Prevention*, vol. 71, pp. 82–92, 2014.
- [17] D. J. Sun, L. Zhang, and F. Chen, "Comparative study on simulation performances of CORSIM and VISSIM for urban street network," *Simulation Modelling Practice and Theory*, vol. 37, pp. 18–29, 2013.
- [18] D. Sun and L. Elefteriadou, "Research and implementation of lane-changing model based on driver behavior," *Transportation Research Record*, no. 2161, pp. 1–10, 2010.
- [19] D. J. Sun and L. Elefteriadou, "Lane changing behavior on urban streets: an 'in-vehicle' field experiment-based study," *Computer-Aided Civil and Infrastructure Engineering*, vol. 27, no. 7, pp. 525–542, 2012.
- [20] Federal Highway Administration (FHWA), *Surrogate Safety Assessment Model and Validation: Final Report FHWA-HRT-08-051*, Federal Highway Administration, Washington, Wash, USA, 2008.
- [21] B. Park and J. Won, *Microscopic Simulation Model Calibration and Validation Handbook: Traffic Operations Laboratory*, Center for Transportation Studies, University of Virginia, Charlottesville, Virginia, Va, USA, 2006.
- [22] Z. Liu, X. Chen, Q. Meng, and I. Kim, "Remote park-and-ride network equilibrium model and its applications," *Transportation Research Part B: Methodological*, vol. 117, pp. 37–62, 2018.
- [23] Z. Liu, S. Wang, W. Chen, and Y. Zheng, "Willingness to board: a novel concept for modeling queuing up passengers," *Transportation Research Part B: Methodological*, vol. 90, pp. 70–82, 2016.

Research Article

Analysis on Port and Maritime Transport System Researches

Lingshan Chen,¹ Xiang Xu ,¹ Ping Zhang ,² and Xiaotian Zhang^{1,3}

¹School of Management, Shanghai University, Shanghai, China

²School of Economic and Management, Tongji University, Shanghai, China

³Oulu Business School, University of Oulu, Oulu, Finland

Correspondence should be addressed to Xiang Xu; xuxiang9623@163.com

Received 30 August 2018; Revised 17 October 2018; Accepted 24 October 2018; Published 21 November 2018

Guest Editor: Zhiyuan Liu

Copyright © 2018 Lingshan 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 examines the past and current research in the container port and maritime field. Using rigorous bibliometric analysis, the paper identifies the core authors/affiliations, their rankings, and collaboration patterns. The analysis of the paper will enable new researchers to quickly build an understanding of the container port and maritime field by reading core authors' papers published in specific journals.

1. Introduction

The maritime industry has made great contributions to the world economy in recent decades. The “Maritime Review by 2015” reported by the United Nations Conference on Trade and Development (UNCTAD) shows that nearly 80% of global commodity trade in volume terms was completed through ports and maritime transport routes. The international maritime transportation industry contributes significantly to the welfare and development of nations adding around \$380 billion a year via freight rates alone to the global economy. At the same time, the total amount of marine transport has steadily increased every year, and in 2014, it reached 9.84 billion tons [1–3]. Standing at the critical interface between inland and sea transportation, the container port is a critical connection between different modes of transportation and represents a critical point in the transportation chain [4]. For a country, maritime transportation not only ensures the import of scarce resources needed for production processes but also facilitates the export of excessive resources, which accumulates more wealth for the country. Maritime transport is also a key to economic globalization [5]. In particular, container transport has become the most important mode of transport in international trade and the new window for the development of foreign economic relationship and trade. Worldwide container port throughput increased from 36 million TEU in 1980 to 614 million TEU

in 2017 and forecasts point to more than 800 million TEU in 2017 [6]. The flourishing industrial growth engendered numerous intellectual problems which, in turn, attracted academic interest. Subsequently, container port and maritime transportation has grown as a unique academic field.

Therefore, it is highly important to develop an overview of this field, which will provide general and historical results that permit a retrospective evaluation. In general, a number of studies have attempted to address conceptualization, methodological issues, theoretical developments, academic taxonomy, and future research directions in areas relevant to maritime logistics literature, such as Lau et al. [7]; Lee and Song [8] survey the extant research in the field of ocean container transport. Shi and Li [5] examine maritime transport researches through a comprehensive review of papers published in 19 transportation journals over the period 2000–2014. Charles [9] provided a global evaluation of the marine academic studies. Woo et al. [10] reviewed published port literature from the 1980s to 2000s in order to investigate the methodological trends in seaport research. Talley [11] reviewed and analyzed maritime transportation academic research, summarized the research topic from maritime journals, and defined future maritime transportation research directions. Chang et al. [12] examined the top 50 authors, 50 affiliations, and 50 countries in the maritime transportation field and discussed the potential correlation between the methodological popularity and author performance.

A number of reviews have been completed on specific aspects of maritime transportation research. For example, Wang et al. [3] pointed out the necessity and importance of port management and operations and suggested more research efforts on potential hot topics. Davarzani et al. [13] examined the past and present research on ‘green ports and maritime logistics’ and identified established research trends and future potential research areas. Several reviews have been more problem-specific and sustainable, such as the review by Mansouri et al. [14], who focused on the use of multiobjective decision methods in sustainable maritime transportation. In addition, with the increasing competition in ports, relevant research also expands rapidly. Sharaf et al. [15] provided an understanding of the efficiency analysis of container port through a comprehensive review of existing studies. Based on empirical evidence, Notteboom [16] analyzed the paths shipping lines and terminal operating companies were walking and also provided an overview of challenges port and maritime companies faced with in an ever-changing competitive environment.

In spite of Chang et al. [12], to the best of the authors’ knowledge, no other studies have analyzed the authors’ affiliations, and countries in maritime research. Our paper differs from Chang et al. [12] in four aspects. First, Chang et al. [12] prespecified a set of journals which are the most closely related to maritime research and they confined the literature search within these journals. In contrast, we search the literature in the whole Web of Science database. Second, different from Chang et al. [12], we conduct a more refined literature search; for example, papers that are related to “maritime” and “fish” are excluded. Third, Chang et al. [12] investigated the whole field of maritime papers and we focus on shipping and container port research. Fourth, Chang et al. [12] judged whether a paper is relevant based on its correlation with seven authors’ papers, which may not be accurate. We judge the relevance of all papers manually. Despite the differences, our research is based on Chang et al. [12]’s seminal work which provided a number of insightful ideas. Using rigorous bibliometric analysis methods, this paper reviews the literature of container port and maritime logistics research to accomplish the following goals. First, we provide some initial statistics of the key journals, authors, and institutions that have contributed to the field. Second, we identify the most active researchers, affiliations, and countries in the container port and maritime field and rank them by different scoring methods. Thus, we provide a better understanding of how maritime transportation research has been undertaken in a quantitative manner. And the ranking also helps to identify active authors. Research of active authors tends to be more advanced. Following their articles can help new comers to obtain new research hotspot more quickly.

The remainder of this paper is organized as follows. Section 2 explains the methodology and details the scope of articles, database searching, scoring methods, and measures of collaborative network. Section 3 reports the ranking of authors and affiliations by all periods of study (1996-2016) and changes over five-year periods (1996-2000, 2001-2005, 2006-2010, and 2011-2017). Section 4 discusses countries/regions’ ranking performance. Section 5 analyzes and

identifies research topics and seminal research areas. Finally, we present the study’s conclusions and discuss the study’s limitations and potential future research directions in Section 6.

2. Methodology

The purpose of the literature review is to map and evaluate the body of the literature and identify potential research gaps. Structured literature reviews are completed by Saunders et al. [17] by iterative using search keywords defined appropriately, searching in the databases, and accomplishing the analysis. Rowley and Slack [18] recommend a structured methodology for scanning resources, designing the mapped structure of the literature review, writing the study, and building the bibliography. Inspired by Rowley and Slack [18] and Seuring and Gold [19], we design a four-step method to collect data and conduct a methodical, comprehensive analysis of the field. We aim to identify the remarkable research, make sure about the classical areas of current research interest, and provide insights for present research and directions for the future.

2.1. Choice of Search Word and Database. Through several trials and errors, we identify suitable search terms and keyword structures. We design the following method to establish the keywords search structure effectively inspired by Rowley and Slack [18] and Soring and Gold [19].

- (i) Build original unit of keywords.
- (ii) Review the search results and make sure whether typical papers and considerable journals are contained in obtained results, and make corresponding modifications to keyword set.
- (iii) Identify the keywords needed to exclude, and make corresponding modifications to keyword set.
- (iv) Search for ‘exclusion research areas’ to confine the search scope, and make corresponding modifications to keyword set.

Initially, we relied on the prior work of Chang et al. [12] in the maritime logistics review papers to define the initial set of keywords. Thirteen search words were suggested based on these previous works. They were “port OR shipping OR maritime OR marine OR terminal OR ship OR liner OR vessel OR seaport OR water transport* OR ocean freight OR container” and “waterway transport*.” First, search words were typed in the Web of Science database. This paper only retrieves academic articles (or journal papers) and rules out conference proceedings, book chapters, dissertations, and theses. The space of journals selected in this paper narrows down to Science Citation Index (SCI), Social Science Citation Index (SSCI), SCI (E), and SSCI-registered ones. Afterwards, we check the resulting articles and journals. A mass of words have a polysemic effect. For example, “vessel” means either a ship or a duct or canal holding or conveying blood or other fluid. The polysemy effect of “vessel” causes the initial search results to include a mass of papers about biological research. We also ruled out the articles that include the following irrelevant words, including “highway OR intersection OR

TABLE 1: Process of the material collection.

		Search results
Step1	Search keywords port OR shipping OR maritime OR marine OR terminal OR ship OR liner OR seaport OR (water transport*) OR (ocean freight) OR container OR (waterway transport*)	81,126
Step2	Exclusion keywords AND NOT highway OR intersection OR helmet OR pedestrian OR fish OR guardrail OR aviation OR airport OR airline OR fishery	78,482
Step3	Remove the irrelevant subject areas	12,279
Step4	Manual refinement of the search results	5,534

helmet OR pedestrian OR fish OR guardrail OR aviation OR airport OR airline” and “fishery.”

The initial search results left about 12279 articles after several trial and error attempts. The search results have expanded to a wide range of thematic areas beyond the scope of this paper. Hence, papers from irrelevant thematic areas need to rule out. Irrelevant subject areas are those that do not belong to the space of container and marine transport system. The unrelated areas were identified through discussions with other senior researchers in the field. The unrelated areas included (1) astronomy, planet sciences, and related areas, (2) agricultural sciences and related areas, (3) medicine, biology, and related areas, (4) physics and related areas, and (5) psychology. Finally, the authors went through 12,279 references and reduced the number of relevant articles to 5,534 as a refinement. Table 1 shows the whole process of the material collection.

2.2. Scoring Method. In this paper, we used two criteria to rank researchers, affiliations, and countries, the number of their publications, and their impact score. The following text describes the composition of each indicator and the rationale for how to use them in detail. The number of publications is an important indicator of academic performance. Numerous studies have used this indicator to approximate author performance [13, 20]. However, only calculating the number of publications has the limitation of measuring the impact of authors on the whole field. In this case, the more crucial part the authors played in an article, the more important evaluation they should be given.

Authors who publish their work in more prestigious journals deserve more respect and higher scores. In this case, it seems appropriate to use the impact factor of a journal to adjust the authors' impact score. There is, in general, a widespread consensus among the authors that the impact factor measures the quality of journals approximately and reasonably. And usage of impact factor is prevalent [21, 22]. The Journal Citation Reports published by Thomson Reuters provided statistical data over the years. A number of journals had no statistical results for certain years, since they recently entered SCI (E) or SSCI, including *Maritime Policy and Management*. In such a case, the impact factor in the entry year was applied to the nonentrance periods. For example, the 2011 impact factor for *Maritime Policy and Management*

was applied to 2000-2010. Afterwards, the impact score was denoted by

$$Score_j = \sum_{y=2000}^{2017} \sum_i \frac{1}{I_j^i} \times IM_y^i \quad (1)$$

where $Score_j$ is the impact score of author j , N is the number of papers published by author j , I_j^i represents the number of corresponding authors in paper i with author j , and IM_y^i is the impact factor of journals that paper i was published in in year y .

Apart from impact score assessment, the number of citations of each paper is also included in the database we have obtained, which is more basic and fundamental and also a good indicator of the author's performance. We measured the total number of citations of an author i by

$$c_i = \sum_{n \in N} ct_n \quad (2)$$

where c_i is the citation score of authors i , N is the total number of publications of author i , and ct_n is the citations of paper n .

3. Analysis of Author and Affiliation

Section 3.1 illustrates the overall trend of publications. Sections 3.2-3.3 report rankings of author and affiliation for 1996-2017, respectively.

3.1. Overall Trend. Figure 1 shows the number of papers published each year. Between 1996 and 2012, the published paper counts in maritime transport system field increased every year. The number peaked in 2012 and decreased thereafter. In 2015, this number began to rise again. The rising paper numbers may be due to the Chinese government vigorously promoting the economic strategy of the Silk Road Economic Belt and the 21st Century Maritime Silk Road. The paper number of Chinese scholars has improved, thereby causing the total number of papers to increase in 2015. Section 3.3 reports the country rankings from 1996 to 2016 and supports this conclusion.

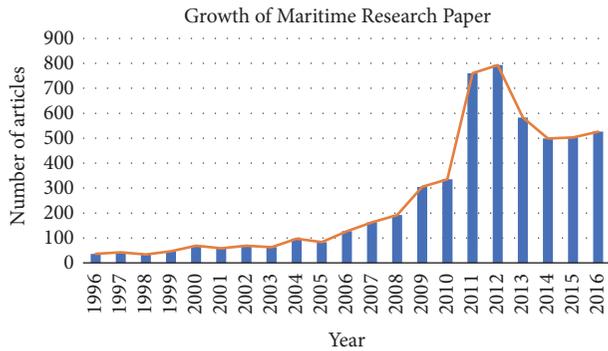


FIGURE 1: Growth of maritime research paper.

3.2. Author Ranking Analysis

3.2.1. Author Ranking for 1996-2016. Table 2 displays the ranking of top 50 authors in the port and maritime transport system field according to the three metrics we mentioned above. Column 3 shows the ranking of the first metric, which is calculated based on the number of papers published. Numbers are counted when the author is one of all coauthors or the single author of a paper. For example, the top ranked Qiang Meng published 61 papers, either as an author or a coauthor. The second metric ranking is the impact score (column 6) that considers the journal impact factor. In terms of paper numbers, Qiang Meng is the top researcher and published 61 papers. Qiang Meng was followed by Shuaian Wang with 52 papers. After those two researchers, the differences between the subsequent authors are not substantial.

Considering the impact score, Meng and Wang still take the first and second places, but their ranks are reversed. The gap between the second and third is relatively large. A notable change can be seen. Among the top 10 authors, Metin Celik, Chung-Yee Lee, Dong-Ping Song, and Lu Zhen emerge. Celik's ranking increased from 16th to 7th. Lee's ranking increased from 24th to 8th. Song's ranking increased from 16th to 9th. Zhen's ranking increased from 35th to 10th.

Significant changes can be observed on the citation score: Nishimura, Etsuko, and Stahlbock, R are the first time appearing in the top 10. Compared with the rank of number of papers, Imai, Akio, Voss, Stefan, and Papadimitriou, Stratos's ranks up more than 25.

3.2.2. Ranking Dynamics of Authors. This section examines the changes in the author's performance over each five-year period. Table 3 shows the dynamics of the impact score ranking of authors. The columns next to the impact score indicate the change of the ranking relative to the previous period in 2001-2005, 2006-2010, and 2011-2015, respectively. Ranking changes of an author will not be shown if he/she was outside the top 100 in the previous period. Overall, the rise and decline of core authors are prominent in the port and maritime transport system field. A notable point is that K.H. Kim maintains a top 10 ranking in every period. Numerous top 50 authors from 1996 to 2000 do not make the rankings in the next period, while most of the top 50 authors in 2001-2005 are new authors. A similar pattern appeared between

2006 and 2010 and 2011 and 2015. This field also had a new strength among the top 5 from 2011 to 2015. A typical example is S.A. Wang. Since all of his articles are published after 2010, he did not appear in the ranking of first three-time periods. Nevertheless, his centrality for 2011-2015 is quite high. Other examples include J.S.L. Lam and L. Zhen. It can be expected that these authors will contribute more to the field in the coming years. Only Notteboom maintains a top 5 position over the entirety of the last decade. The changeable ranking means this field is full of competition and opportunity.

Another point worth noting is that the five-year impact scores of active authors have a significant raise. For example, the impact scores of the top 3 between period 3 and period 4 soared from 9.05, 8.73, and 7.79, respectively, to 34.45, 31.03, and 17.98, respectively. It may have been influenced by the growing number of publications and more publications being published in high-impact journals (impact score = $\sum_{y=2000}^{2017} \sum_i^N (1/I_i^y) \times IM_y^i$).

3.2.3. Core Authors and Collaboration Patterns. Derek Price prompted the celebrated "square root law" that states that half of the scientific papers are contributed by the top square root of the total number of scientific authors. The law was first proposed in Little Science-Big Science [23] and is heuristically based on Lotka's inverse square law. The Price's law is calculated from the following equation:

$$M = 0.749 \sqrt{N_{max}} \quad (3)$$

where N_{max} is the maximum number of the articles by one author. M is the minimum number of articles by a core author, which means the author whose published paper number is above M is the core author. In our data sample, N_{max} equals 61, and according to Price's law M equals 5.8. Thus, there are more than 190 authors who have published more than 6 papers in the field.

For further analysis, we also tried to analyze the cooperation pattern of the core authors. The cooperation pattern means different authors appearing in multiple joint articles frequently. Table 4 shows the results. An interesting observation is that the core authors (such as S.A. Wang, Q. Meng, E.P. Chew, L.H. Lee, and K. Fagerholt) in the container port and maritime field do frequently cooperate. This result indicates that most container port and maritime researchers are willing to cooperate with their familiar partners.

The top articles were considered as the lead articles of a research area [13]. The titles of core authors' top 10 articles in the container port and maritime research area are also shown in Table 5. All of the top 3 articles are the overview in the port and maritime research area which may help new comers to understand quickly about container port and maritime field. The rest of the top 10 articles are interested in specific topics, including economics, operations research, business, and planning.

3.3. Affiliation Ranking Analysis

3.3.1. Affiliation Ranking for 2000-2015. Table 6 displays affiliation rankings for 1996-2016. In paper counts, the National

TABLE 2: Rankings by author.

Rank	Author	No. papers	Rank	Author	Impact Score	Rank	Author	Citation Score
1	Meng, Qiang	61	1	Wang, Shuaian	47.95	1	Kim, Kap Hwan	1713
2	Wang, Shuaian	52	2	Meng, Qiang	41.66	2	Imai, Akio	1214
3	Kim, Kap Hwan	47	3	Notteboom, Theo	25.63	3	Christiansen, Marielle	1189
4	Fagerholt, Kjetil	45	4	Lam, Jasmine Siu Lee	25.00	4	Fagerholt, Kjetil	1147
5	Lam, Jasmine Siu Lee	36	5	Kim, Kap Hwan	23.63	5	Meng, Qiang	1107
5	Notteboom, Theo	36	6	Fagerholt, Kjetil	23.06	6	Papadimitriou, Stratos	1083
7	Wang, Jin	31	7	Celik, Metin	20.41	7	Voss, Stefan	1030
8	Christiansen, Marielle	30	8	Lee, Chung-Yee	19.00	8	Nishimura, Etsuko	1021
9	Lee, Loo Hay	28	9	Song, Dong-Ping	17.44	9	Stahlbock, R	960
10	Parola, Francesco	26	10	Zhen, Lu	17.35	10	Wang, Shuaian	751
10	Lu, Chin-Shan	26	11	Monios, Jason	17.10	11	Cullinane, Kevin	744
10	Chew, Ek Peng	26	12	Ding, Ji-Feng	15.76	12	Laporte, Gilbert	680
10	Lee, Paul Tae-Woo	26	13	Ng, ManWo	14.41	13	Kozan, Erhan	659
14	Luo, Meifeng	25	14	Lu, Chin-Shan	13.99	14	Song, Dong-Wook	624
15	Lun, Y. H. Venus	24	15	Talley, Wayne K	13.45	15	Vis, Iris F. A.	614
16	Celik, Metin	23	16	Tovar, Beatriz	13.27	16	Notteboom, Theo	577
16	Lee, Der-Horng	23	17	Lee, Paul Tae-Woo	12.62	17	Meisel, Frank	562
16	Song, Dong-Ping	23	18	Thai, Vinh V	12.56	18	Park, YM	540
19	Cheng, T. C. Edwin	22	19	Ducruet, Cesar	12.56	19	Wang, Jin	522
19	Cullinane, Kevin	22	20	Christiansen, Marielle	12.15	20	Liu, JY	519
22	Ng, Adolf K. Y.	21	21	Lun, Y. H. Venus	11.75	21	Ronen, David	506
21	Thai, Vinh V	21	22	Lee, Loo Hay	11.65	22	Wan, YW	505
21	Yang, Zaili	21	23	Luo, Meifeng	11.61	23	Steenken, D	500
24	Chang, Young-Tae	20	24	Petering, Matthew E. H.	11.22	24	Bierwirth, Christian	484
24	Lee, Chung-Yee	20	25	Wilmsmeier, Gordon	11.18	25	Lee, Der-Horng	450
24	Li, Kevin X	20	26	Lee, Der-Horng	11.14	26	Lu, Chin-Shan	436
27	Lai, Kee-hung	19	27	Lai, Kee-hung	10.96	27	Tongzon, J	431
27	Laporte, Gilbert	19	28	Akyuz, Emre	10.91	28	de Koster, R	424
29	Huynh, Nathan	18	29	Cullinane, Kevin	10.76	29	Lee, Loo Hay	422
29	Papadimitriou, Stratos	18	30	Parola, Francesco	10.54	30	Linn, RJ	422
31	Ducruet, Cesar	17	31	Wang, Jin	10.17	31	Celik, Metin	411
31	Imai, Akio	17	32	Boysen, Nils	10.09	32	Andersson, Henrik	395
31	Tovar, Beatriz	17	33	Chou, Chien-Chang	10.08	33	Legato, Pasquale	391
31	Yip, Tsz Leung	17	34	Dong, Jing-Xin	9.88	34	Kujala, Pentti	389
35	Hu, Hao	16	35	Wang, Xinchang	9.83	35	Chew, Ek Peng	385
35	Kavussanos, Manolis G	16	36	Meisel, Frank	9.62	36	Song, Dong-Ping	375
35	Lim, Andrew	16	37	Ng, Adolf K. Y.	9.52	37	Murty, Katta G.	368
35	Monios, Jason	16	38	Chew, Ek Peng	9.47	38	Lam, Jasmine Siu Lee	360
35	Pisinger, David	16	39	Laporte, Gilbert	9.47	39	Ng, WC	320
35	Song, Dong-Wook	16	40	Yang, Yi-Chih	8.98	40	Wang, Teng-Fei	317
35	Talley, Wayne K	16	41	Yang, Zaili	8.90	41	Kavussanos, Manolis G.	312
35	Yang, Zhongzhen	16	42	Chang, Young-Tae	8.81	42	Zhang, CQ	291
35	Zhen, Lu	16	43	Kujala, Pentti	8.57	43	Hummels, David	286
44	Ferrari, Claudio	15	44	Psaraftis, Harilaos N.	8.56	44	Pisinger, David	282
44	Hvattum, Lars Magnus	15	45	Cheng, T. C. Edwin	8.44	45	Gue, KR	270
44	Kozan, Erhan	15	46	Cariou, Pierre	8.24	46	Lun, Y. H. Venus	269
44	Kujala, Pentti	15	47	Yip, Tsz Leung	8.20	47	Ng, Adolf K. Y.	249
44	Voss, Stefan	15	48	Pisinger, David	8.13	48	Cheng, T. C. Edwin	240
44	Wilmsmeier, Gordon	15	49	Sun, Zhuo	8.09	49	Er, I. Deha	237
44	Wilson, William W	15	50	Bell, Michael G. H.	7.90	50	Hasle, Geir	233

TABLE 3: Ranking dynamics of authors.

	Author	Impact Score (1996-2000)	Author	Impact Score (2001-2005)	Author	Impact Score (2006-2010)	Author	Impact Score (2011-2015)	
1	Kim, Kap Hwan	3.82	Kim, Kap Hwan	3.75	-	Celik, Metin	-	Wang, Shuaian	34.45
2	Bennett, P	2.44	Suryanata, K	1.65	▲	Chou, Chien-Chang	-	Meng, Qiang	31.03
3	Peterson, K	1.52	Peterson, K	1.52	-	Notteboom, Theo	▲	Notteboom, Theo	17.98
4	Lei, J	1.45	Lei, J	1.45	-	Er, I. Deha	▲	Lam, Jasmine Situ Lee	14.10
5	Zhou, JJ	1.45	Zhou, JJ	1.45	-	Kim, Kap Hwan	▼	Fagerholt, Kjetil	12.40
6	Hartmann, Soenke	1.44	Hartmann, Soenke	1.44	-	Ding, Ji-Feng	-	Song, Dong-Ping	11.55
7	Hemp, P	1.37	Imai, Akio	1.41	▲	Lai, Kee-hung	-	Zhen, Lu	9.98
8	Matthews, HS	1.35	Nishimura, Etsuko	1.41	▲	Petering, Matthew E. H.	▲	Lee, Chung-Yee	9.66
9	Hart, R	1.23	Papadimitriou, Stratos	1.41	▲	Lee, Der-Horng	▲	Kim, Kap Hwan	9.38
10	Wang, Jin	1.08	Lillie, N	1.38	▲	Roso, Violeta	-	Lee, Loo Hay	8.85
11	Holden, H	1.00	Hemp, P	1.37	▼	Imai, Akio	▲	Akyuz, Emre	7.63
12	Gierloff-Emden, HGR	0.96	Matthews, HS	1.35	▼	Song, Dong-Wook	-	Dong, Jing-Xin	7.31
13	van Driel, Hugo	0.93	Oakley, Susan	1.29	-	Vis, Iris F. A.	-	Ng, ManWo	7.17
14	Suryanata, K	0.89	Hart, R	1.23	▼	Hummels, David	-	Yang, Zaili	7.05
15	Umemoto, KN	0.89	Karlsen, A	1.17	-	Chen, Jiang Hang	-	Kujala, Pentti	6.98
16	Krasnopolsky, VM	0.89	Cullinane, Kevin	1.15	▲	Roe, Michael	-	Tovar, Beatriz	6.94
17	Schiller, H	0.89	Wang, Jin	1.13	▼	Lu, Chin-Shan	▼	Ducruet, Cesar	6.74
18	Lillie, N	0.88	Ng, WC	1.08	-	Goodchild, Anne V	-	Chew, Ek Peng	6.57
19	Zhang, CQ	0.85	Holden, H	1.00	▼	Ducruet, Cesar	-	Wilmsmeier, Gordon	6.56
20	Jetlund, AS	0.84	Gierloff-Emden, HGR	0.96	▼	Papadimitriou, Stratos	▼	Ng, Adolf K. Y.	6.42
21	Karimi, IA	0.84	van Driel, Hugo	0.93	▼	Hirashima, Yoichi	-	Celik, Metin	6.28
22	Fossen, Thor I.	0.82	Umemoto, KN	0.89	▼	Fremont, Antoine	-	Liu, Zhiyuan	6.21

TABLE 3: Continued.

Author	Impact Score (1996-2000)	Author	Impact Score (2001-2005)	Author	Impact Score (2006-2010)	Author	Impact Score (2011-2015)
23 Park, YM	0.82	Liu, Jiyin	0.89	Demaria, Federico	2.75	Yang, Yi-Chih	5.99
24 Ronen, D.	0.80	Wan, Yat-wah	0.89	Fagerholt, Kjetil	2.74	Cariou, Pierre	5.85
25 Jacobson, A	0.80	Krasnopolsky, VM	0.89	Kozan, Erhan	2.69	Monios, Jason	5.69
26 Farrell, JA	0.79	Schiller, H	0.89	Asteris, Michael	2.67	Rudd, Murray A.	5.68
27 Li, W	0.79	Tongzon, Jose	0.85	Meng, Qiang	2.67	Talley, Wayne K.	5.67
28 Lee, RM	0.78	Zhang, CQ	0.85	Paulauskas, Vytautas	2.67	Sun, Zhuo	5.65
29 Linn, RJ	0.76	Legato, Pasquale	0.84	Meisel, Frank	2.63	Laporte, Gilbert	5.65
30 Imai, Akio	0.76	Jetlund, AS	0.84	Ng, Adolf K. Y.	2.61	Pisinger, David	5.56
31 Nishimura, Etsuko	0.76	Karimi, IA	0.84	Vasiliauskas, Aidas Vasilis	2.55	Psaraffis, Harilaos N.	5.52
32 Papadimitriou, Stratos	0.76	Ronen, D.	0.80	Sharma, Mithun J.	2.53	Petering, Matthew E. H.	5.43
33 Ausloos, M	0.76	Jacobson, A	0.80	Yu, Song Jin	2.53	Lee, Paul Tae-Woo	5.37
34 Ivanova, K	0.76	Lu, Chin-Shan	0.79	Nishimura, Etsuko	2.51	Goerlandt, Floris	5.34
35 Fagerholt, Kjetil	0.75	Farrell, JA	0.79	Lee, Yusin	2.50	Wang, Hua	5.24
36 Fairlie, DP	0.73	Li, W	0.79	Konings, Rob	2.47	Hvattum, Lars Magnus	4.98
37 Lucke, AJ	0.73	Donderi, DC	0.78	Saeed, Naima	2.43	Yip, Tsz Leung	4.92
38 Singh, Y	0.73	McFadden, S	0.78	Merrick, Jason R. W.	2.43	Christiansen, Marielle	4.91
39 Tyndall, JDA	0.73	Lee, RM	0.78	Ramirez-Marquez, Jose Emmanuel	2.33	Hanninen, Maria	4.89
40 Fung, Michael K.	0.72	Bish, EK	0.77	Kahraman, Cengiz	2.29	Wang, Jin	4.83

TABLE 3: Continued.

	Author	Impact Score (1996-2000)	Author	Impact Score (2001-2005)	Author	Impact Score (2006-2010)	Author	Impact Score (2011-2015)
41	Bish, EK	0.67	Linn, RJ	0.76	Hastings, Justin V.	2.27	Featherstone, David	4.82
42	Yun, JE	0.67	Ausloos, M	0.76	Cebi, Selcuk	2.18	Lee, Der-Horng	4.77
43	Marin, PL	0.66	Ivanova, K	0.76	Panayides, Photis M	2.18	Hu, Zhi-Hua	4.77
44	Sicotte, R	0.66	Umemoto, K	0.76	Goulielmos, Alexander M.	2.11	Kontovas, Christos A.	4.76
45	Legato, Pasquale	0.66	Sciomachen, Anna	0.75	Cao, Jin Xin	2.10	Knapp, Sabine	4.52
46	Bhargava, HK	0.66	Koutsavdis, E	0.74	Park, Changkyu	2.09	Luo, Meifeng	4.51
47	Snoap, KJ	0.66	Rhee, SH	0.74	Seo, Junyong	2.09	Nathan Huynh	4.50
48	Altiok, T	0.64	Fairlie, DP	0.73	Wu, Wei-Ming	2.07	Yeo, Gi-Tae	4.43
49	Cullinane, Kevin	0.63	Lucke, AJ	0.73	Fossen, Thor I.	2.05	Meisel, Frank	4.41
50	Gold, C	0.62	Singh, Y	0.73	Low, Joyce M. W	2.03	Bell, Michael G. H.	4.40

TABLE 4: The most prolific paired authors.

Rank	Paired authors	Number of joint publications	Rank	Paired authors	Number of joint publications
1	Meng, Q, Wang, SA	34	11	Papadimitriou, S, Nishimura, E	12
2	Chew, EP, Lee, LH	25	12	Fagerholt, K, Christiansen, M	12
3	Lai, KH, Cheng, TCE	17	13	Tan, KC, Chew, EP	11
4	Cheng, TCE, Lun, YHV	15	14	Tan, KC, Lee, LH	11
5	Yang, ZL, Wang, J	15	15	Wang, SA, Liu, ZY	11
6	Boile, M, Theofanis, S	14	16	Monios, J, Wilmsmeier, G	10
7	Lai, KH, Lun, YHV	14	17	Parola, F, Satta, G	10
8	Lee, DH, Chen, JH	13	18	Lee, DH, Cao, JX	10
9	Papadimitriou, S, Imai, A	13	19	Lai, KH, Wong, CWY	9
10	Imai, A, Nishimura, E	12	20	Christiansen, M, Andersson, H	9

TABLE 5: The top 10 articles of port and maritime research area.

Author	Title	Citation
Steenken, D; Voss, S; Stahlbock, R	Container terminal operation and operations research - a classification and literature review	500
Stahlbock, Robert; Voss, Stefan	Operations research at container terminals: a literature update	423
Vis, IFA; de Koster, R	Transshipment of containers at a container terminal: An overview	290
Christiansen, M; Fagerholt, K; Ronen, D	Ship routing and scheduling: Status and perspectives	270
Hummels, David	Transportation costs and international trade in the second era of globalization	242
Bierwirth, Christian; Meisel, Frank	A survey of berth allocation and quay crane scheduling problems in container terminals	228
Imai, A; Nishimura, E; Papadimitriou, S	The dynamic berth allocation problem for a container port	222
Kim, KH; Park, YM	A crane scheduling method for port container terminals	200
Tongzon, J; Heng, W	Port privatization, efficiency and competitiveness: Some empirical evidence from container ports (terminals)	169
Cullinane, K; Wang, TF; Song, DW; Ji, P	The technical efficiency of container ports: Comparing data envelopment analysis and stochastic frontier analysis	168

University of Singapore (NUS), Hong Kong Polytechnic University (HKPU), Norwegian University of Science and Technology, Delft University of Technology, and the University of Antwerp (UA) are in the top 5. The top schools are all strong regardless of how they are ranked. The most productive school, NUS, published 184 papers and was followed by HKPU with 174 papers. As with the authors' ranking, the gap between the second and third in the affiliation rankings is also relatively large (60 papers). Below these schools, the gap between the subsequent affiliations is no longer substantial. The top ten affiliations for 1996-2016 published 986 papers, which equate to an annual average of 46.87 papers. The best performing schools is still NUS. Another notable point is the sudden increase of Nanyang Technological University (NTU) in the ranking. As for the rank of citation score, it is stable

and similar to the rank of impact score except for some slight fluctuations. But there is still some sudden increase that needs to be noteworthy. Univ Georgia Inst Technol's rank ups 16 places to 9th. Univ Hamburg is not among the top 50 in the impact score but places to 5th in citation score.

The affiliation ranking shows a similar pattern to author rankings. NUS have an outstanding container port and maritime researcher; Meng and HKPU have Wang. The two authors publish in impact score journals in container port and maritime research. It seems to indicate a direct relationship between research infrastructure and researchers. The size of the research infrastructure and the number of researchers available in the field may impact the researchers' effectiveness and strength of knowledge conversion. In other words, the more the numerous researchers and knowledge, the higher

TABLE 6: Rankings by affiliations.

Rank	School	No. Paper	Rank	School	Impact Score	Rank	School	Citation Score
1	Natl Univ Singapore	184	1	Natl Univ Singapore	148.06	1	Natl Univ Singapore	3342
2	Hong Kong Polytech Univ	174	2	Hong Kong Polytech Univ	130.64	2	Norwegian Univ Sci & Technol	2525
3	Norwegian Univ Sci & Technol	114	3	Norwegian Univ Sci & Technol	69.67	3	Pusan Natl Univ	2179
4	Delft Univ Technol	93	4	Delft Univ Technol	64.03	4	Hong Kong Polytech Univ	1558
5	Univ Antwerp	84	5	Nanyang Technol Univ	59.83	5	Univ Hamburg	1075
6	Erasmus Univ	82	6	Erasmus Univ	58.04	6	Erasmus Univ	1052
7	Nanyang Technol Univ	73	7	Pusan Natl Univ	56.32	7	Istanbul Tech Univ	951
8	Shanghai Jiao Tong Univ	68	8	Dalian Maritime Univ	53.49	8	Univ Antwerp	934
9	Pusan Natl Univ	66	9	Istanbul Tech Univ	52.56	9	Georgia Inst Technol	868
10	Dalian Maritime Univ	62	10	Tech Univ Denmark	50.78	10	Hong Kong Univ Sci & Technol	802
11	Univ Genoa	61	11	Univ Antwerp	50.53	11	Univ Calif Berkeley	785
12	Univ Piraeus	58	12	Natl Taiwan Ocean Univ	47.26	12	Natl Cheng Kung Univ	777
13	Natl Taiwan Ocean Univ	55	13	Univ Genoa	46.33	13	Tech Univ Denmark	758
14	Istanbul Tech Univ	54	14	Hong Kong Univ Sci & Technol	43.11	14	Delft Univ Technol	719
15	Hong Kong Univ Sci & Technol	53	15	Natl Cheng Kung Univ	42.08	15	HEC Montreal	699
16	Tech Univ Denmark	52	16	Natl Kaohsiung Marine Univ	41.71	16	Kobe Univ	689
17	Univ Plymouth	51	17	Univ Plymouth	41.24	17	Univ Plymouth	645
18	Natl Cheng Kung Univ	47	18	Old Dominion Univ	39.36	18	Univ Hong Kong	635
18	Rutgers State Univ	47	19	Univ Piraeus	38.91	19	Univ Calabria	618
20	Cardiff Univ	43	20	Shanghai Jiao Tong Univ	38.59	20	Nanyang Technol Univ	574
21	Georgia Inst Technol	40	21	Cardiff Univ	32.64	21	Liverpool John Moores Univ	569
22	Natl Kaohsiung Marine Univ	39	22	Natl Tech Univ Athens	31.14	22	Norwegian Marine Technol Res Inst MARINTEK	559

TABLE 6: Continued.

Rank	School	No. Paper	Rank	School	Impact Score	Rank	School	Citation Score
23	Univ Rijeka	38	23	Vilnius Gediminas Tech Univ	31.00	23	Univ Genoa	559
24	Edinburgh Napier Univ	37	24	Edinburgh Napier Univ	30.94	24	Univ Halle Wittenberg	539
24	Liverpool John Moores Univ	37	25	Georgia Inst Technol	29.72	25	Univ Wisconsin	525
24	Natl Tech Univ Athens	37	26	Shanghai Univ	29.51	26	Univ London Imperial Coll Sci Technol & Med	513
27	Old Dominion Univ	36	27	Aalto Univ	29.35	27	Rutgers State Univ	492
27	Texas A&M Univ	36	28	Liverpool John Moores Univ	27.55	28	Univ Michigan	488
29	Univ Aegean	35	29	Inha Univ	26.71	29	Shanghai Jiao Tong Univ	461
30	Inha Univ	34	30	Univ Las Palmas Gran Canaria	26.42	30	Aalto Univ	458
30	Norwegian Marine Technol Res Inst MARINTEK	34	31	Univ Hong Kong	26.02	31	Univ Piraeus	421
30	Univ Hong Kong	34	32	Univ Calabria	24.92	32	Univ Elect Sci & Technol China	416
33	Univ Calif Berkeley	32	33	Norwegian Marine Technol Res Inst MARINTEK	24.05	33	Chalmers University of Technology	411
34	Chung Ang Univ	31	34	Univ Wollongong	24.02	34	Cardiff Univ	409
34	Univ Valencia	31	35	Shanghai Maritime Univ	23.97	35	Natl Tech Univ Athens	409
36	Shanghai Maritime Univ	30	36	Univ Aegean	23.03	36	Univ Tecn Lisboa	369
36	Shanghai Univ	30	37	Univ Belgrade	21.50	37	Queensland Univ Technol	363
38	HEC Montreal	29	38	Queensland Univ Technol	21.50	38	Kobe Univ Mercantile Marine	354
38	Univ Calabria	29	39	Univ Wisconsin	20.74	39	Univ Newcastle Upon Tyne	346
38	Univ Illinois	29	40	Chalmers	20.68	40	Univ Maryland	344
41	Eindhoven Univ Technol	28	41	Chang Jung Christian Univ	20.53	41	City Univ Hong Kong	342

TABLE 6: Continued.

Rank	School	No. Paper	Rank	School	Impact Score	Rank	School	Citation Score
41	MIT	28	42	Eindhoven Univ Technol	20.27	42	Dalian Maritime Univ	336
41	Univ Las Palmas Gran Canaria	28	43	Texas A&M Univ	20.09	43	Molde Univ Coll	336
44	Aalto Univ	27	44	Wuhan Univ Technol	19.05	44	Univ Las Palmas Gran Canaria	334
45	Univ Hamburg	26	45	Rutgers State Univ	18.60	45	Univ Southampton	332
45	Univ Oviedo	26	46	Chung Ang Univ	18.59	46	Natl Kaohsiung Marine Univ	329
45	Univ Wisconsin	26	47	Univ Illinois	18.49	47	Purdue Univ	324
45	Wuhan Univ Technol	26	48	Univ London Imperial Coll Sci Technol & Med	18.29	48	Antwerp Maritime Acad	318
49	Chalmers	25	49	Univ Seville	18.04	49	Chinese Acad Sci	305
49	City Univ Hong Kong	25	50	Univ Valencia	18.03	50	Ecole Polytech	303

the probability of the affiliation obtaining substantial knowledge conversion.

3.3.2. Ranking Dynamics for Affiliations. In Table 7, the change of dynamic ranking of affiliations is relatively small. HKPU, NUS, and EU are consistently high. As we mentioned above, there seems to be a direct relationship between the affiliations and researchers. According to Foray and Lundvall [24], human capital (including graduates, highly skilled personnel, and public and private researchers) somewhat tends to flow to have more of the benefits from positive spillovers. Conversely, if there is not too much of a brain drain, the dynamics of affiliation ranking will remain relatively stable. However, if affiliations seek to remain competitive, it will have to retain a positive welfare system and take care of its producers in order to stem a brain drain.

3.3.3. Collaboration Patterns. We also analyzed the cooperation pattern of affiliations. Table 8 shows the top ten paired affiliations. The result indicates that affiliations do not stay in static cliques. They are willing to cooperate with new partners in order to achieve diversity and novelty. Core affiliations in the maritime field play the role of spreading advanced research results and promoting the development of the field.

4. Analysis of the Countries/Regions

Table 9 displays country rankings from 1996 to 2016. USA, China, and England are the top 3 countries irrespective of the scoring method. Although none of the authors or affiliations

in the USA reached the top five, it still has dominant positions in these fields. It is true that local academic researchers could contribute to the competitiveness of the territory. However, in the context of world metropolis, the competition is no longer subject to natural geographical constraints, but more dependent on the territory itself to attract and retain human capital. The advantage of country competitiveness is actively built and not passively suffered. This may also indicate that the American government is good at attracting and promoting research.

Going further into country publications in the port and maritime transport system field, the top 20 countries ranking by the annual number of publications are discussed. Figure 2 shows the change among the top 20 countries over time. It can be seen that the general trend of annual journal publication has increased. Over the past decade, however, the publications of China have grown much faster than in the United States and other countries. In addition, China has kept up with the United States approximately with the same number of journal publications in the port and maritime transport system field since 2014, as indicated in Figure 2. Interestingly, the pioneers are not American or Chinese, even though they are the most influential countries in this field, since they did not publish papers from the beginning.

5. Analysis of Research Terms

In bibliometric analysis, it is useful to select several keywords as a representation of important research topics in this field if a researcher wants to explore every facet of a field's major

TABLE 7: Ranking dynamics of affiliations.

	school	impact score (1996-2000)	school	impact score (2001-2005)	school	impact score (2006-2010)	school	impact score (2011-2015)
1	Pusan Natl Univ	6.44	Pusan Natl Univ	7.12	Natl Univ Singapore	34.97	Natl Univ Singapore	98.16
2	Natl Univ Singapore	3.68	Natl Univ Singapore	4.33	Istanbul Tech Univ	24.40	Hong Kong Polytech Univ	42.30
3	Chinese Acad Sci Univ	3.14	Hong Kong Polytech Univ	3.33	Hong Kong Polytech Univ	19.92	Norwegian Univ Sci & Technol	39.96
4	Queensland Univ	2.93	Univ Hawaii Manoa	3.30	Pusan Natl Univ	18.40	Nanyang Technol Univ	33.29
5	Hong Kong Polytech Univ	2.73	Chinese Acad Sci Univ	3.14	Natl Kaohsiung Marine Univ	12.50	Dalian Maritime Univ	32.20
6	Univ Edinburgh	2.44	Univ Queensland	2.93	Univ Calabria	12.16	Pusan Natl Univ	31.76
7	Penn State Univ	2.39	Univ Hong Kong	2.60	Univ Genoa	10.31	Delft Univ Technol	29.53
8	Norwegian Univ Sci & Technol	2.36	Norwegian Univ Sci & Technol	2.56	Erasmus Univ	10.01	Univ Antwerp	29.44
9	Erasmus Univ	2.28	Erasmus Univ	2.52	Georgia Inst Technol	9.90	Erasmus Univ	28.32
10	Carnegie Mellon Univ	2.25	Univ Genoa	2.52	Univ Wisconsin	9.85	Shanghai Maritime Univ	27.10
11	Hong Kong Univ Sci & Technol	1.89	Univ Texas	2.44	Univ Hong Kong	9.64	Tech Univ Denmark	25.07
12	Univ Hawaii Manoa	1.78	Hong Kong Univ Sci & Technol	2.29	Norwegian Univ Sci & Technol	9.54	Old Dominion Univ	24.22
13	Univ So Calif	1.76	Carnegie Mellon Univ	2.25	Univ Piraeus	8.96	Univ Wollongong	22.55
14	World Bank	1.71	Univ Piraeus	2.02	Univ Antwerp	8.41	Univ Genoa	22.01
15	Univ Tennessee	1.52	Natl Tech Univ Athens	1.99	Natl Cheng Kung Univ	8.39	Aalto Univ	21.88
16	Univ Genoa	1.51	Natl Cheng Kung Univ	1.83	Delft Univ Technol	8.06	Univ Plymouth	21.64
17	Univ Calif Riverside	1.40	Univ So Calif	1.76	Univ Plymouth	7.70	Shanghai Univ	19.62
18	Georgia Inst Technol	1.38	World Bank	1.71	Shanghai Maritime Univ	7.62	Shanghai Jiao Tong Univ	19.36
19	Natl Tech Univ Athens	1.36	Univ Calabria	1.68	Univ So Calif	7.29	Natl Kaohsiung Marine Univ	18.79
20	Univ Calabria	1.32	Queensland Univ Technol	1.67	Vrije Univ Amsterdam	6.75	Natl Taiwan Ocean Univ	18.51

TABLE 7: Continued.

	school	impact score (1996-2000)	school	impact score (2001-2005)	school	impact score (2006-2010)	school	impact score (2011-2015)
21	Florida Int Univ	1.28	Georgia Inst Technol	1.61	Chinese Acad Sci	6.66	Hong Kong Univ Sci & Technol	18.46
22	Univ Piraeus	1.25	Florida Int Univ	1.60	Chalmers Univ	6.46	Istanbul Tech Univ	17.41
23	USN	1.24	Univ Tennessee	1.52	Washington Chang Jung Christian Univ	6.16	Cardiff Univ	16.12
24	Swedish Univ Agr Sci	1.23	Fluent Inc	1.48	Cardiff Univ	5.86	Liverpool John Moores Univ	15.03
25	Kobe Univ	1.19	Univ Calif Riverside	1.40	Cardiff Univ	5.46	Tongji Univ	13.87
26	Liverpool John Moores Univ	1.17	Penn State Univ	1.32	Shanghai Jiao Tong Univ	5.19	Norwegian Marine Technol Res Inst	13.81
27	Univ Hong Kong	1.03	Univ Adelaide	1.29	Univ Western Australia	5.17	MARINTEK	13.19
28	Univ Liege	1.00	Liverpool John Moores Univ	1.27	Univ London Imperial Coll Sci Technol & Med	5.14	Edinburgh Napier Univ	13.16
29	Calif State Univ Bakersfield	0.96	Univ Missouri	1.25	Vilnius Gediminas Tech Univ	5.10	Univ La Laguna	13.12
30	Univ Sydney	0.93	Swedish Univ Agr Sci	1.23	Rutgers State Univ	5.10	Natl Tech Univ Athens	13.06
31	Rutgers State Univ	0.92	Kobe Univ	1.19	Univ Portsmouth	5.08	Univ Las Palmas Gran Canaria	12.88
32	Univ Glamorgan	0.92	World Maritime Univ	1.16	Natl Taiwan Ocean Univ	5.04	Imperial Coll Sci Technol & Med	12.08
33	Univ Ottawa	0.90	Natl Taiwan Ocean Univ	1.05	Queensland Univ Technol	4.69	Univ Calabria	12.06
34	Korea Adv Inst Sci & Technol	0.90	Texas A&M Univ	1.04	Kobe Univ	4.63	Univ Aegean	11.41

TABLE 7: Continued.

	school	impact score (1996-2000)	school	impact score (2001-2005)	school	impact score (2006-2010)	school	impact score (2011-2015)
35	Univ Missouri GKSS	0.90	Univ Delaware	1.02	Nanyang Technol Univ	4.29	ExxonMobil Res & Engrn Co	11.34
36	Forschungszen- trum Geesthacht GmbH	0.89	Univ Liege	1.00	Univ Halle Wittenberg	4.24	Eindhoven Univ Technol	11.33
37	NOAA	0.89	Newcastle Univ	0.98	Univ Politecn Cataluna	4.19	Chung Ang Univ	11.11
38	Cornell Univ	0.88	Calif State Univ Bakersfield	0.96	Natl Tech Univ Athens	4.19	Univ Seville	11.07
39	Univ Delaware	0.88	Univ Sydney	0.93	Univ Ulsan	4.18	Inha Univ	10.97
40	Univ Illinois	0.81	Kobe Univ	0.91	Scottish Agr Coll	4.13	Univ Piraeus	10.66
41	Univ E Anglia	0.80	Univ Ottawa	0.90	Univ N Carolina	4.10	Univ S Carolina	10.35
42	Queensland Univ Technol	0.76	Univ Michigan	0.89	Molde Univ Coll	3.96	Univ British Columbia	10.18
43	Aplus Flash Technol Inc	0.75	GKSS Forschungszen- trum Geesthacht GmbH	0.89	Dalian Maritime Univ	3.82	Univ Bergen	10.18
44	Kyungnung Univ	0.74	NOAA	0.89	Univ Calif Berkeley	3.77	Newcastle Univ	9.76
45	Univ Carlos III Madrid	0.73	Cornell Univ	0.88	Stevens Inst Technol	3.76	Univ Sydney	9.50
46	Natl Cheng Kung Univ	0.73	Fern Univ Hagen	0.82	CNR	3.56	Georgia Inst Technol	9.47
47	Univ Pittsburgh	0.72	Univ Illinois	0.81	Hanyang Univ	3.48	Univ Valencia	9.27
48	Korea Maritime Inst	0.72	Univ E Anglia	0.80	Purdue Univ	3.43	Univ Politecn Valencia	9.08
49	Tech Univ Berlin	0.72	McGill Univ	0.78	Tsinghua Univ	3.36	Wuhan Univ Technol	9.04
50	Univ Wales Coll Cardiff	0.72	CNRS ULP	0.78	Univ Sci & Technol China	3.29	Univ Wisconsin	8.81

TABLE 8: The most prolific paired affiliations.

Rank	Paired affiliations	Number of joint publications
1	Norwegian Marine Technol Res Inst MARINTEK, Norwegian Univ Sci & Technol	27
2	Univ Naples Parthenope, Univ Genoa	19
3	Natl Univ Singapore, Univ Wollongong	14
4	Antwerp Maritime Acad, Univ Antwerp	14
5	HEC Montreal, Norwegian Univ Sci & Technol	11
6	Nanyang Technol Univ, RMIT Univ	11
7	Hong Kong Polytech Univ, Shanghai Univ	11
8	Hong Kong Polytech Univ, Chung Ang Univ	8
9	Nanyang Technol Univ, Univ Antwerp	8
10	Dalian Maritime Univ, Univ Antwerp	8

research topics and their relationships down to the finest detail. Term Frequency-Inverse Document Frequency (TF-IDF) promoted by Salton and Buckley [25] is a typical method of identifying important terms by combining their popularity and discrimination. The TF-IDF method can also be applied to bibliometric analysis. For example, Jaboska-Sabuka et al. [26] used TF-IDF to identify informative words from publication keywords of the research domain in order to predict research trends. Roche [27] used TF-IDF to select publication keywords of scientific fields and categorized them into unusual terms, established terms, and cross-sectional terms. $(tf - idf_t)$ is calculated for each word. Salton et al. [28] from the following equation:

$$tfidf_t = \sum_d tf_{t,d} \times \log \frac{N}{df_t} \quad (4)$$

where $tf_{t,d}$ is frequency of the word t in the document d , N is number of the articles, and df_t is the number of articles where word t existed.

5.1. Overall Analysis. Various research terms are observed in articles published in the time period from 1996 to 2016. Table 10 shows the top 60 research terms of these fields. From Table 10, we can indicate that port and maritime transport system field have been classified in the literature in terms of shipping or port research and their respective methodologies applied in the research. We also built the word cooccurrence table to create highly cooccurring word sets. The input words in the table satisfy the following two limitations. First, the $tf-idf$ factor of these words is greater than the A threshold value. Second, the number of publications which contained a candidate word is within the specified range by the B threshold. An analysis of Table 11 reveals that the specific shipping topics include seafarers, short sea shipping, shipping performance/management, shipping finance, and shipping safety. Specific port topics include port governance/privatization, port performance, port state control, port competition, and port choice.

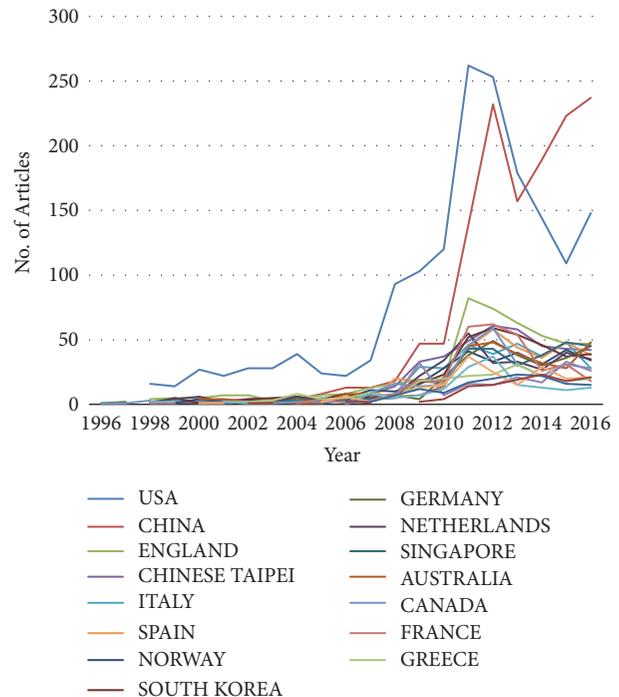


FIGURE 2: Trend of the countries/regions.

Table 12 shows the dynamics of the research terms. Although there are several new words that arise throughout time, the rise and fall of research terms are prominent in the port and maritime transport system field. In 2006-2010, some new words such as berth and vessel turned up as new research objects. In 2001-2015, the word emission turned up as a new research topic and reached the top 5. This may be caused by some environment protection policy such as Regulation 14 of the IMO that required ships to switch to low sulfur fuels in ECA areas. Another notable point is the rank change of the words risk and liner. Risk was a top 8 word in 1996-2000, but fell to 18th in 2011-2015. By inspecting

TABLE 9: The ranking of the countries/regions.

Rank	Country	No. papers	Rank	Country	Impact Score
1	USA	993	1	USA	1170.25
2	PEOPLES R CHINA	714	2	PEOPLES R CHINA	972.35
3	ENGLAND	305	3	ENGLAND	384.71
4	CHINESE TAIPEI	263	4	CHINESE TAIPEI	354.65
5	SOUTH KOREA	205	5	SPAIN	269.14
6	ITALY	196	6	NORWAY	265.79
7	SPAIN	194	7	SINGAPORE	263.14
8	NORWAY	189	8	ITALY	263.10
9	GERMANY	183	9	NETHERLANDS	246.82
10	NETHERLANDS	182	10	GERMANY	246.18
11	SINGAPORE	172	11	SOUTH KOREA	244.03
12	AUSTRALIA	154	12	AUSTRALIA	212.31
13	CANADA	141	13	FRANCE	212.21
14	FRANCE	141	14	CANADA	197.57
15	GREECE	131	15	TURKEY	166.71
16	TURKEY	107	16	GREECE	161.82
17	JAPAN	99	17	BELGIUM	117.86
18	BELGIUM	87	18	DENMARK	113.96
19	SWEDEN	78	19	SWEDEN	108.00
20	DENMARK	74	20	JAPAN	107.91
21	CROATIA	69	21	PORTUGAL	76.12
23	PORTUGAL	56	22	INDIA	69.68
24	SCOTLAND	52	23	SCOTLAND	67.00
25	IRAN	51	24	FINLAND	64.27
26	INDIA	49	25	IRAN	61.00
27	RUSSIA	49	26	BRAZIL	58.16
28	POLAND	40	27	SWITZERLAND	44.00
29	BRAZIL	39	29	WALES	40.62
30	FINLAND	38	30	POLAND	39.64
31	WALES	36	31	CROATIA	36.24
32	LITHUANIA	33	32	LITHUANIA	34.05
33	SWITZERLAND	26	33	SERBIA	27.93
34	CHILE	24	34	ISRAEL	24.99
35	ISRAEL	23	35	CYPRUS	21.23
36	SLOVENIA	23	36	NEW ZEALAND	21.16
37	SERBIA	20	37	RUSSIA	21.15
38	MALAYSIA	18	38	AUSTRIA	20.00
39	NEW ZEALAND	16	39	CHILE	19.75
40	SOUTH AFRICA	16	40	MALAYSIA	19.25
41	AUSTRIA	14	41	SLOVENIA	17.52
42	CYPRUS	12	42	U ARAB EMIRATES	16.55
43	MEXICO	12	43	SOUTH AFRICA	14.76
44	U ARAB EMIRATES	12	44	MEXICO	13.88
45	UKRAINE	11	45	THAILAND	12.11
46	MONTENEGRO	10	46	LEBANON	10.98
47	IRELAND	8	47	IRELAND	9.59
48	LEBANON	8	48	CZECH REPUBLIC	9.18
49	THAILAND	8	49	NIGERIA	6.83
50	ARGENTINA	6	50	MONTENEGRO	6.49

TABLE 10: The top 60 research terms.

Word	TF-IDF	Word	TF-IDF
port	201.99	management	58.21
container	171.73	simulation	57.80
terminal	130.71	supply	57.59
shipping	121.37	dynamic	56.89
ship	120.68	liner	56.88
cost	100.25	company	56.76
transport	97.69	logistics	56.56
control	92.95	cargo	56.49
network	90.09	factor	56.44
service	86.66	impact	56.28
algorithm	85.17	route	55.70
maritime	82.58	rate	55.62
vessel	77.78	planning	55.03
market	75.80	development	54.87
transportation	74.55	optimization	54.57
risk	73.53	safety	54.32
freight	72.68	yard	54.17
approach	72.35	truck	53.48
operation	71.54	price	52.32
optimal	70.83	trade	52.11
data	67.46	traffic	51.59
efficiency	64.82	heuristic	51.58
crane	64.69	marine	51.10
scheduling	63.60	capacity	51.04
process	63.14	level	50.22
policy	61.66	economic	50.06
strategy	60.64	flow	48.94
emission	59.64	sea	48.66
chain	59.52	condition	47.73
industry	58.51	function	47.45

TABLE 11: The top 6 words cooccurrence table for port and shipping.

Cooccurrence word	Frequent	Cooccurrence word	Frequent
shipping management	603	port perform	824
shipping perform	354	port state control	383
shipping network	232	port governance/government	209
shipping short-term	168	port choice	188
shipping safety	157	port competition	183
shipping finance	91	port private(privatization)	135

the papers that include the word risk, we determined that the author Jin Wang published numerous papers examining maritime risk in 1996-2000, but he did not publish as much from 2011 to 2015, and his rank fell from 10 to 40. We can draw

an obvious conclusion that Shuaian Wang is a specialized researcher investigating liners, and he is the top author in 2011-2015. This finding indicates that a specific research may be influenced by one or two authors.

TABLE 12: Ranking dynamics of research terms.

	Word	TF-IDF (1996-2000)	Word	TF-IDF (2001-2005)		Word	TF-IDF (2006-2010)		Word	TF-IDF (2011-2015)	
1	freight	2.88	container	3.46	▲	port	16.37	▲	port	20.67	-
2	network	2.34	port	3.24	▲	container	14.64	▼	container	18.52	-
3	science	2.11	yard	2.45	-	shipping	8.27	▲	shipping	14.50	-
4	container	2.09	cost	2.28	▲	terminal	8.07	▲	ship	13.86	▲
5	carrier	2.00	truck	2.19	▲	cost	7.86	▼	emission	12.10	-
6	intermodal	1.99	heuristic	1.98	-	maritime	7.45	-	terminal	11.02	▼
7	transportation	1.83	ship	1.90	▲	transport	7.42	▲	cost	9.75	▼
8	risk	1.82	terminal	1.86	▲	market	7.09	▲	network	9.46	▲
9	port	1.76	vehicle	1.76	-	crane	6.94	▲	transport	9.29	▼
10	policy	1.65	safety	1.76	▲	transportation	6.85	▲	service	8.69	▲
11	terminal	1.56	policy	1.73	▼	yard	6.27	▼	freight	8.48	▼
12	truck	1.56	transport	1.71	▲	berth	6.17	-	control	8.47	-
13	ocean	1.54	crane	1.67	-	scheduling	5.97	-	algorithm	8.43	▲
14	service	1.51	transportation	1.65	▼	ship	5.94	▲	approach	8.16	▲
15	market	1.50	algorithm	1.62	-	network	5.87	▼	liner	7.93	-
16	operation	1.49	capacity	1.61	-	operation	5.79	-	crane	7.91	▼
17	ship	1.45	market	1.57	▼	service	5.61	▲	maritime	7.73	▼
18	transport	1.44	shipping	1.48	-	approach	5.43	-	risk	7.66	▼
19	cost	1.41	service	1.48	▼	vessel	5.36	-	optimal	7.61	-
20	safety	1.31	product	1.44	-	algorithm	5.34	-	cargo	7.48	-

6. Conclusions

In this paper, we have analyzed maritime-related academic research. We utilized bibliometric analysis method to illustrate the evolution of this field. The two criteria used in ranking were the number of published papers and the impact score (reflecting the prestige of the journals). We focused on the papers published in journals included in the Science Citation Index and Social Science Citation Index.

Ranking criteria influences the overall rankings for authors and affiliations sensitively. In terms of the impact score, the most active researchers are S.A. Wang, Q. Meng, T. Notteboom, J.S.L. Lam, and K.H. Kim. The most active affiliations were the National University of Singapore, Hong Kong Polytechnic University, the Norwegian University of Science and Technology, Delft University of Technology, and the Nanyang Technology University. Affiliations rankings seem to have been affected by authors ranking. For example, the top affiliation has the top author. However, the network features of authors and affiliations are quite different. At the microlevel, overall research terms in the container port and maritime field are identified by the TF-IDF algorithm.

This paper has limitations. In terms of the scoring method, especially impact score, the impact factor does not reflect a journals quality perfectly. For instance, several highly respected journals have a low impact factor. There are several other indicators that can measure the impact, such as the H index and the number of citations for authors.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they are no conflicts of interest regarding the publication of this paper.

References

- [1] L. Zhen, "Tactical berth allocation under uncertainty," *European Journal of Operational Research*, vol. 247, no. 3, pp. 928–944, 2015.
- [2] L. Zhen, "Modeling of yard congestion and optimization of yard template in container ports," *Transportation Research Part B: Methodological*, vol. 90, pp. 83–104, 2016.
- [3] L. Zhen, E. P. Chew, and L. H. Lee, "An integrated model for berth template and yard template planning in transshipment hubs," *Transportation Science*, vol. 45, no. 4, pp. 483–504, 2011.
- [4] C. Zhou, L. I. Haobin, W. Wang, L. H. Lee, and E. P. Chew, "Connecting the belt and road through sea-rail collaboration," *Frontiers of Engineering Management*, vol. 4, no. 3, pp. 315–324, 2017.
- [5] W. Shi and K. X. Li, "Themes and tools of maritime transport research during 2000-2014," *Maritime Policy & Management*, vol. 44, no. 2, pp. 151–169, 2017.

- [6] *Annual Review of Global Container Terminal Operators*, Drewry Shipping Consultants, London, 2016.
- [7] Y.-Y. Lau, C. Ducruet, A. K. Y. Ng, and X. Fu, "Across the waves: a bibliometric analysis of container shipping research since the 1960s," *Maritime Policy & Management*, vol. 44, no. 6, pp. 667–684, 2017.
- [8] C.-Y. Lee and D.-P. Song, "Ocean container transport in global supply chains: Overview and research opportunities," *Transportation Research Part B: Methodological*, vol. 95, pp. 442–474, 2017.
- [9] K. Charles, "Marine science and blue growth: Assessing the marine academic production of 123 cities and territories worldwide," *Marine Policy*, vol. 84, pp. 119–129, 2017.
- [10] S.-H. Woo, S. J. Pettit, D.-W. Kwak, and A. K. C. Beresford, "Sea-port research: A structured literature review on methodological issues since the 1980s," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 7, pp. 667–685, 2011.
- [11] W. K. Talley, "Maritime transportation research: Topics and methodologies," *Maritime Policy & Management*, vol. 40, no. 7, pp. 709–725, 2013.
- [12] Y. T. Chang, K. S. Choi, A. Jo, and H. Park, "Top 50 authors, affiliations, and countries in maritime research," in *Proceedings of the and countries in maritime research. International Journal of Shipping Transport Logistics*, vol. 10, p. 7, 2018.
- [13] H. Davarzani, B. Fahimnia, M. Bell, and J. Sarkis, "Greening ports and maritime logistics: A review," *Transportation Research Part D: Transport and Environment*, vol. 48, pp. 473–487, 2015.
- [14] S. A. Mansouri, H. Lee, and O. Aluko, "Multi-objective decision support to enhance environmental sustainability in maritime shipping: a review and future directions," *Transportation Research Part E: Logistics and Transportation Review*, vol. 78, pp. 3–18, 2015.
- [15] E. Sharaf, M. Z. Shah, and A. S. Al-Iraqi, "Efficiency analysis of analysis of container port: a review," in *Proceedings of Seatuc Symposium*, 2014.
- [16] T. E. Notteboom, "Container Shipping And Ports: An Overview," *Review of Network Economics*, vol. 3, no. 2, 2010.
- [17] M. Saunders, P. Lewis, and A. Thornhill, "Research methods for business students (5th edn)," *Qualitative Market Research*, vol. 3, no. 4, pp. 215–218, 2009.
- [18] J. Rowley and F. Slack, "Conducting a literature review," *Management Research News*, vol. 27, no. 6, pp. 31–39, 2004.
- [19] S. Seuring and S. Gold, "Conducting content-analysis based literature reviews in supply chain management," *Supply Chain Management Review*, vol. 17, no. 5, pp. 544–555, 2012.
- [20] C. R. Carter, D. B. Vellenga, J. J. Gentry, and B. J. Allen, "Affiliation of authors in transportation and logistics academic journals: A reassessment," *Transportation Journal*, vol. 44, no. 2, pp. 54–64, 2005.
- [21] C. Hoeffel, "journal impact factors," *Allergy*, vol. 53, no. 12, pp. 1225–1225, 1998.
- [22] D. C. Greenwood, "Reliability of journal impact factor rankings," *BMC Medical Research Methodology*, vol. 7, 2007.
- [23] D. J. De Solla Price, "Networks of scientific papers," *Science*, vol. 149, no. 3683, p. 510, 1965.
- [24] D. Foray and B. A. Lundvall, "The knowledge-based economy," *Modeled Measured Simulated Boca Raton Fl Universal*, vol. 50, no. 5, 1996.
- [25] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.
- [26] M. Jaboska-Sabuka, R. Sitarz, and A. Kraslawski, "Forecasting research trends using population dynamics model with Burgers' type interaction," *Journal of Informetrics*, vol. 8, no. 1, pp. 111–122, 2014.
- [27] I. Roche, D. Besagni, C. François, M. Hörlesberger, and E. Schiebel, "Identification and characterisation of technological topics in the field of Molecular Biology," *Scientometrics*, vol. 82, no. 3, pp. 663–676, 2010.
- [28] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1974.

Research Article

Application of Finite Mixture of Logistic Regression for Heterogeneous Merging Behavior Analysis

Gen Li 

School of Transportation, Southeast University, Nanjing 210096, China

Correspondence should be addressed to Gen Li; gilg4226307@aliyun.com

Received 4 July 2018; Revised 15 October 2018; Accepted 4 November 2018; Published 21 November 2018

Guest Editor: Lele Zhang

Copyright © 2018 Gen Li. 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 finite mixture of logistic regression model (FMLR) was applied to analyze the heterogeneity within the merging driver population. This model can automatically provide useful hidden information about the characteristics of the driver population. EM algorithm and Newton-Raphson algorithm were used to estimate the parameters. To accomplish the objective of this study, the FMLR model was applied to a trajectory dataset extracted from the NGSIM dataset and a 2-component FMLR model was identified. The important findings can be summarized as follows: The studied drivers can be classified into two components. One is called Risk-Rejecting Drivers. These drivers are consistent with previous studies and primarily merge in as soon as possible and have a distinct preference for the large gaps. The other is the Risk-Taking Drivers that are much less sensitive to the gap size and pay more attention to surrounding traffic conditions such as the speed of front vehicle in the auxiliary lane and lead space gap between the merging vehicle and its leading vehicles in the auxiliary lane. Risk-Taking Drivers use the auxiliary lane to get to the further downstream or less congested area of the main lane. The proposed model can also produce more precise predicting accuracy than logistic regression model.

1. Introduction

Congestion has become one of the most serious economic and social problems and has drawn great attention from the public, transportation research scientists, transportation managers, and so on. Understanding the causes and mechanism of traffic congestion can help traffic managers formulate targeted policies to make better use of the existing transportation infrastructures.

Merging areas are the bottleneck of freeway. Merging behavior is one of the typical mandatory lane changes when vehicles have to move from an on-ramp to the main road. It has been claimed in some studies that merging behavior at merging areas affects traffic operations and may trigger traffic congestions and breakdowns [1, 2]. Thus it is important to analyze the merging behaviors to help understand the mechanism of traffic jams to some extent from a microscopic viewpoint and build more accurate traffic simulation models.

Recently, driver heterogeneity has drawn great attention in microscopic traffic flow studies. Several studies investigated the driver heterogeneity during car-following process

[3–6]. Accommodating heterogeneity within the driver population is important in building microscopic traffic models. To investigate the heterogeneity in merging behaviors, a finite mixture of logistic regression (FMLR) model was proposed in this paper. This model can incorporate the unobserved heterogeneity and automatically segments the merging drivers into different homogeneous populations. More specifically, this paper aims to achieve the following objectives:

- (i) Prove the existence of heterogeneity among merging drivers.
- (ii) Identify different driving styles and attitudes during merging process.
- (iii) Model the merging behavior more accurately.

The present study is organized as follows. The next section will provide a critical review on the existing relevant literature followed by Section 3, which describes the NGSIM data used in this paper. Section 4 gives the methodology to build FMLR model. Results and discussions are presented in Section 5.

Finally, the conclusions and future work are presented in Section 6.

2. Literature Review

Several methods have been adopted to model merging behavior, among which gap acceptance theory was the most widely used method [8–13]. The most important assumption in gap acceptance theory was that a driver makes a lane change when both the lead and lag gaps in the target lane are larger than the so-called critical gap. The critical gap is determined by the characteristics of the drivers, traffic conditions, and so on [14]. Gap acceptance models were initially built to estimate the capacity of unsignalized intersections. Different distributions of critical gaps were assumed in various studies [15–17]. Gipps [18] first used the gap acceptance theory to propose a comprehensive framework of lane-changing model. Gipps's framework has been widely used in several merge models [19, 20] and microscopic traffic simulation software [21, 22]. Different definitions of critical gap were used in these models and software.

Gap acceptance theory was often criticized as its basic assumption is often inconsistent with the real world observation because some lane change behaviors occurred when only the lead or lag gap or even none of them are larger than the critical gap [14, 23, 24]. To overcome this deficiency, discrete choice models such as binary logit model were used by some researchers [14, 25–27]. Built upon a series of studies [9, 10, 28], a framework for merging behavior with latent plans was introduced by Choudhury *et al.* [29]. Normal merge, merge with courtesy, and forced merge were considered in this framework. However, Marczak *et al.* [14] pointed out that in this framework only accepted gaps were considered and rejected gaps were ignored; and some of the estimated coefficients in the model were not significant.

Traffic behaviors are always uncertain and variable and heterogeneity cannot be ignored in traffic studies. Some studies investigated the heterogeneity among the macroscopic traffic flow [30, 31]. Others studied the heterogeneity in car following behaviors from microscopic viewpoint by deriving the joint distribution of model coefficients depending on an empirical basis [4, 5, 32–34]. However, only a few studies were found to investigate driver heterogeneity in lane changing models. A two-step clustering analysis was proposed by Li and Sun [35] to analyze heterogeneity of the merging maneuvers. However, this study ignored the heterogeneity during gap selection and decision process. An empirical analysis conducted by Daamen *et al.* [23] showed that different merging strategies might be adopted by different drivers under different traffic conditions. It has been pointed out by Keyvan-Ekbatani *et al.* [36] that different strategies might be used during gap selection process; however the sample size was too small to perform statistically relevant tests and build merging model.

Thus, a FMLR model was introduced in this paper to model the gap selection behaviors during merging process and investigate the heterogeneity among merging drivers. The FMLR model takes the advantage of two techniques: clustering and regression analysis. The model naturally incorporates

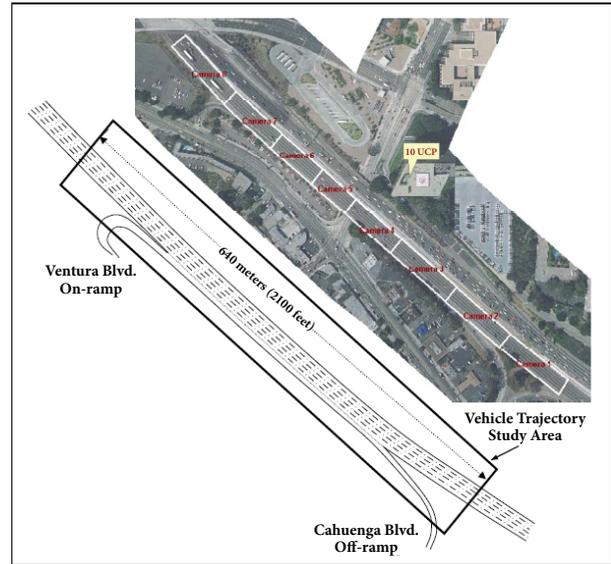


FIGURE 1: The section of US 101 [7].

the unobserved heterogeneity into logistic regression model and automatically segments the drivers into different homogeneous populations. The proposed FMLR model can explain the different strategies in merging behaviors.

3. Data Preparation

The NGISM dataset has been widely used for traffic flow and traffic simulation studies and proved to have high accuracy. Thus, in this paper, the vehicle trajectory data in NGISM dataset collected on a segment of southbound U.S. Highway 101 (Hollywood Freeway) in Los Angeles, CA, are chosen [37]. Figure 1 shows the site for U.S. Highway 101. This US-101 section is 640 meters long and has five main lanes and one auxiliary lane. The vehicle trajectories were collected from 7:50 a.m. to 8:35 a.m. on June 15, 2005. The road section was covered by eight cameras and the dataset was updated at a resolution of 10 frames per second [7]. The dataset has three data subsets, all of which were collected in 15 minutes.

In this study, we focus on the behavior of merging vehicles and only trajectory data in the weaving section were used. However, it has been pointed out that the original trajectory data contain some noise and errors, which are caused by the system errors and tracking errors [38–41]. Several methods have been proposed to filter the data [38–40] or re-extract the trajectory data [41]. Re-extracting can produce the most accurate data especially the acceleration data, which however would also make too much effort. In this paper, a smoothing method called sEMA developed by Thiemann *et al.* [38] is applied to reduce the noise and errors. The sEMA method is also adopted in other studies of merging behaviors and has been proved to be able to provide enough precision for lane change studies [42–44]. This data smoothing technique was applied as follows:

(1) The velocities and accelerations of vehicles are directly estimated from the longitudinal positions.

TABLE 1: Examples with the same global coordinates in the first and second subsets.

Data Point	Sub dataset 1				Sub dataset 2			
	Vehicle ID	Frame ID	Local x	Local y	Vehicle ID	Frame ID	Local x	Local y
1	33	424	54.612	1397.746	36	847	54.612	1438.019
2	33	429	54.687	1420.332	1070	4878	54.687	1460.518
2	63	290	67.936	514.811	1472	5857	67.936	550.085

TABLE 2: Examples with the same global coordinates in the first and third subsets.

Data Point	Sub dataset 1				Sub dataset 3			
	Vehicle ID	Frame ID	Local x	Local y	Vehicle ID	Frame ID	Local x	Local y
1	42	446	53.395	1449.048	1721	8609	53.395	1483.814
2	63	483	41.056	1494.004	1280	6744	41.056	1528.773
3	296	967	53.8340	1389.247	905	4719	53.834	1424.013

(2) The locations (both local lateral and longitudinal coordinates), velocities, and accelerations of vehicles are smoothed by the symmetric exponential moving average filter (sEMA) proposed by Thiemann *et al.* [38] to decrease measurement errors in the data. The smoothing times of sEMA method are set as the suggested values for the U.S. Highway 101 dataset in Thiemann *et al.* [38].

Although the random errors can be reduced by the smoothing process, there are still some errors in the data. Thus, the following heuristic rules are applied to filter the datasets:

- (1) Filter out the trajectories when there are no putative leading vehicles or putative following vehicles on the adjacent main lane. Such trajectories are recorded at the beginning or ending of the video tape and cannot provide the interactions of merging vehicles with their surrounding vehicles.
- (2) Filter out the trajectories when putative leading or putative following vehicle of a merging vehicle runs around the lane boundary (it keeps touching the lane boundary before lane change or turns back the original lane in about 1 second). These trajectories are always caused by the tracking errors.

After filtering, a searching process was conducted to check the consistency of the local coordinates and global coordinates. Linear regression was performed between local coordinates and global coordinates for each subdataset. Three linear relationships were obtained for each subset:

$$Localy_1 = 0.3209globalx_1 - 1.1326globaly_1 \quad (1)$$

$$Localy_2 = 0.3291globalx_2 - 1.1334globaly_2 \quad (2)$$

$$Localy_3 = 0.3209globalx_3 - 1.1333globaly_3 \quad (3)$$

R^2 of three linear relationships are 0.9996, 0.9997, and 0.9997, respectively. It means that the local y of three subsets in US-101 datasets are inconsistent with each other. We cannot find simple linear relationship between local x and global x in US-101 dataset. This could be caused by the specific

coordinate system used and the special geometric shape of the road sections. It also could be caused by measuring errors.

To further verify the inconsistency of the US-101 dataset, several data points that have the same global coordinates among the three subsets were searched and obtained. By checking the local coordinates (local x and local y), it was found that the three subsets of US-101 dataset are consistent in local x, but inconsistent in local y. Tables 1 and 2 show the examples having the same global coordinates in the first and second subsets and in the first and third subsets.

One can find that, for the points with the same global coordinates, the three subdatasets have the same local x, but different local y. In the local longitudinal coordinate, the upstream edge (0 m) in datasets 1 is at 12.275m in dataset 2 and 10.598 m in dataset 3. Thus, the three datasets must be unified by using the local coordinates of one of the three subsets.

At every instant when offered a new gap, a merging vehicle driver assesses traffic conditions to decide whether to accept the offered gap or not. One merging vehicle could only accept one gap but could reject several gaps. After data processing, trajectories of 374 merging vehicles consisting of 925 observations were extracted from the dataset. The explanatory variables that may affect a driver's merging decision used as candidates for analyzing the merging behavior model are shown in Table 3.

4. Methodology

4.1. Finite Mixture of Logistic Regression. The FMLR model is based on the idea that the observed data come from a population with several subpopulations or components [45, 46]. The overall population is modeled as a mixture of the groups using finite mixture models.

Let \mathbf{X} and \mathbf{Y} denote random vectors with N samples and each sample has M_n observations $(\mathbf{x}_i, \mathbf{y}_i)$ ($i = 1, \dots, M_n, n = 1, \dots, N$). Here, the response vector \mathbf{Y} has values in \mathbb{R}^d and the explanatory vector \mathbf{X} has values in \mathbb{R}^p . Then, a FMLR with K components has the form

$$h(\mathbf{y} | \mathbf{x}, \psi) = \sum_{k=1}^K \pi_k f(\mathbf{y} | \mathbf{x}, \theta_k) \quad (4)$$

TABLE 3: Descriptions of the explanatory variables.

Variable	Descriptions
D_n^i (m)	The size of the i^{th} offered gap of merging vehicle n
V_n^i (m/s)	The speed of merging vehicle n at i^{th} offered gap.
Y_n^i (m)	The longitudinal position of the merging vehicle n to the start of the auxiliary lane.
ΔV_{nPL}^i (m/s)	The speed difference between the putative leading vehicle and the merging vehicle n at offered gap i .
ΔV_{nPF}^i (m/s)	The speed difference between the putative following vehicle and the merging vehicle n at offered gap i .
δ	Existence of a lead vehicle in the merge lane. If there is a lead vehicle in the merge lane, $\delta = 1$; otherwise, $\delta = 0$.
ΔD_{nlg}^i (m)	Lead gap of merging vehicle n in the auxiliary lane at offered gap i .
V_{nLead}^i (m)	The speed of the leading vehicle in the auxiliary lane at offered gap i .
ΔV_{nLead}^i (m/s)	The speed difference between the leading vehicle in the auxiliary lane and the merging vehicle n at offered gap i .

$$\sum_{k=1}^K \pi_k = 1, \quad \pi_k > 0 \quad (5)$$

where $h(\mathbf{y} | \mathbf{x}, \Psi)$ is the conditional density of \mathbf{y} given \mathbf{x} and θ_k , π_k is the mixing proportion, θ_k is the component-specific parameter vector for the density function f , and $\psi = (\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K)$ is the vector of all parameters.

Several finite mixture models can be extended based on (4) and (5). For multivariate normal f and $\mathbf{x} \equiv \mathbf{1}$ we get a finite mixture of Gaussians without a regression part, also known as model-based clustering. If f is a univariate normal density with component-specific mean $\beta_k' \mathbf{x}$ and variance σ_k^2 , we have $\theta_k = (\beta_k', \sigma_k^2)$, and (4) describes a finite mixture of linear regression, also called latent class linear regression model or cluster-wise regression [47]. If f is a member of the exponential family, we get a FMLR models [48, 49].

The analyst does not observe directly which component, $k = 1, \dots, K$, generated observation \mathbf{y}_i . The model assumes that individuals are distributed heterogeneously with a discrete distribution within the population. In order to impose the constraints in (2), the mixing proportions are parameterized with a multinomial logit form [50, 51]:

$$\pi_k = \frac{\exp(\alpha_k)}{\sum_{k=1}^K \exp(\alpha_k)}, \quad \alpha_K = 0 \quad (6)$$

The constraint on α_K is imposed because only $K - 1$ parameters are needed to specify. The last proportion is one minus the sum of the first $K - 1$.

If individual specific characteristics are provided, the mixing proportions are extended as [50, 51]

$$\pi_{ik} = \frac{\exp(\theta_k \mathbf{z}_i)}{\sum_{k=1}^K \exp(\theta_k \mathbf{z}_i)}, \quad \theta_K = 0 \quad (7)$$

where θ_k is the vector of component-specific parameters and \mathbf{z}_i is an optional set of individual-specific characteristics for observation i .

For the observed random sample, $(\mathbf{x}_i, \mathbf{y}_i)$ ($i = 1, \dots, N$), the log likelihood function for ψ is given by

$$\log L(\psi) = \sum_{n=1}^N \sum_{i=1}^{M_n} \log h(\mathbf{y}_i | \mathbf{x}_n, \psi) \quad (8)$$

$$= \sum_{i=1}^N \sum_{n=1}^{M_n} \log \left(\sum_{k=1}^K \pi_k h(\mathbf{y}_i | \mathbf{x}_n, \theta_k) \right)$$

The maximum likelihood (ML) estimate of ψ is given by an appropriate root of the likelihood equation,

$$\frac{\partial \log L(\psi)}{\partial \psi} = 0 \quad (9)$$

The conditional probability that observation $(\mathbf{x}_i, \mathbf{y}_i)$ belongs to component j is given by

$$P(j | \mathbf{x}_i, \mathbf{y}_i, \psi) = \frac{\pi_j f(\mathbf{y}_i | \mathbf{x}_i, \theta_j)}{\sum_{k=1}^K \pi_k f(\mathbf{y}_i | \mathbf{x}_i, \theta_k)} \quad (10)$$

The conditional probabilities can be used to segment data by assigning each observation to the component with maximum conditional probability [50, 51]. A probabilistic segmentation of the data into K components can be obtained in terms of the fitted conditional probabilities. In the FMLR model we consider the latent component-indicator variables $\hat{z}_n = \hat{z}_{n1}, \dots, \hat{z}_{nK}$, $n = 1, \dots, N$, to classify each single observation:

$$z_{nk} = \begin{cases} 1, & \text{if } y_n \text{ belongs to component } k \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

The estimator of z_{jk} , \hat{z}_{nk} is

$$\hat{z}_{nk} = \begin{cases} 1, & \text{if } \hat{\pi}_k(\mathbf{y}_n; \hat{\Psi}) \geq \hat{\pi}_h(\mathbf{y}_n; \hat{\Psi}), (h = 1, \dots, K; h \neq k) \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

4.2. Model Parameter Estimation. Parameters of FMLR models can be efficiently estimated through the EM algorithm [52].

(1) *Initialization Step.* Start with an initial seed (guess) for the parameter $\hat{\psi}$ using the K-means clustering algorithm [53].

(2) *E-Step.* Estimate the conditional component probabilities, $\hat{\pi}_{ik}$, for each observation using (7) and derive the mixing proportions as

$$\hat{\pi}_k = \frac{1}{N} \sum_{i=1}^N \hat{\pi}_{ik} \quad (13)$$

(3) *M-Step.* Maximize the log-likelihood for each component separately using the conditional probabilities as weights:

$$\max_{\theta_k} \sum_{i=1}^N \hat{\pi}_{ik} \log(y_i | \mathbf{x}_i, \theta_k) \quad (14)$$

The EM algorithm alternates between the expectation and the maximization steps until the likelihood improvement falls under a prespecified threshold or a maximum number of iterations are reached.

But the drawbacks of EM algorithm are its possible slow convergence rate and long processing time in computer. Thus, in this paper, Latent GOLD 5.0 is used to estimate the parameters. Latent GOLD 5.0 can take the advantages of both EM and Newton-Raphson algorithms. It first uses EM algorithm to get close to the final solution and then switches to Newton-Raphson to finish estimation [54].

The most important and difficult step in building FMLR model is to determine K , the number of components. Since K is not a parameter, hypotheses on K cannot be tested directly. BIC or AIC [50, 51, 55, 56] are generally used as criterion to determine K . In this study, we determined K based on BIC:

$$BIC_{\text{model}} = -2LL + \gamma \log(N) \quad (15)$$

where LL is the log-likelihood value, γ is the number of free parameters to be estimated, and N is the number of observations in the data. A lower BIC value indicates a better model.

5. Results and Discussion

5.1. Results. To select an optimal model, we apply the FMLR model having an increasing number of components from 1 to 4 to fit, and apply Bayesian Information Criterion (BIC) as the indicator to select the most appropriate number of components. Table 4 shows the BIC values of models for different number of components. It can be observed from Table 4 that the lowest BIC value occurs at $K = 2$. Hence, it is plausible to select $K = 2$ as a proper number of components.

To select the model variables, the forward-selection method is adopted in this paper. It starts with no variables in the model, tests the addition of each variable using Wald-statistics, and adds the variable that gives the most statistically significant improvement of the fit. In this paper, variables will be added one by one until none produce a significant Wald-statistic in all components.

Table 5 shows the estimation results. For comparison, the result of logistic regression is also provided. In this paper, the

TABLE 4: BIC value of FMLR model.

The Number of Components	BIC Value
$K = 1$	790.2955
$K = 2$	773.3871
$K = 3$	808.3826
$K = 4$	843.3061

component mixing proportions are a set of fixed constants (see (6)), as no sociodemographic characteristics of drivers are available in this dataset. The proportion of merging vehicle drivers in each component as indicated by H value in Table 5 is 67.2% and 32.8%, respectively.

By using (10)-(12), 374 drivers are classified into two components. One is the larger component, comprising 298 drivers and 612 observations, and the other is the smaller component, containing 75 drivers and 314 observations. To better understand the classification results, the mean values and standard deviations of related attribute variables are shown in Table 6.

5.2. Discussion. As seen from significance levels of parameters of Component 1 in Table 5, ΔD_{nlg}^i and ΔV_{nLead}^i fail to be significant at the 99% level. These suggest that front vehicles in auxiliary lane do not alter drivers' merge decisions in this component. Another impressive characteristic of this component is that the drivers have a distinct preference for the larger gaps. The negative sign of V_n^i indicates that drivers in this component tend to decrease their speeds during merging process. Consistent with previous studies, the decrease of speed difference between merging vehicle and putative leading vehicle and a gap located further towards the end of the auxiliary lane also increase the probability of accepting the current gap.

It is interesting to find that the parameter of D_n^i in Component 2 is much smaller than that in Component 1, which means the drivers in Component 2 do not pursue larger gaps as drivers in Component 1. In addition, speed difference between merging vehicle and putative leading vehicle is still important during merging process. Different from Component 1, ΔD_{nlg}^i and V_{nLead}^i are considered by drivers in Component 2. The sign of the parameter for ΔD_{nlg}^i is positive, suggesting that space in the auxiliary lane also affects the merging behaviors of drivers in Component 2 and the merging vehicle has a high probability of accepting a gap when there is an adequate space in front of the merging vehicle. One interesting finding from Table 5 is that the sign of the parameter for V_{nLead}^i is negative, suggesting that drivers in Component 2 are more likely to delay merge when the leading vehicle moves too fast. One possible reason for this result might be that when the leading vehicles move faster in the auxiliary lane, the drivers are provided more space in the auxiliary lane and they are using the auxiliary lane to reach further downstream in the main lane.

As illustrated in Table 6, the related variables show obvious differences across the two components. The average numbers of rejected gaps of the two components are 1.05

TABLE 5: Model estimation results of FMLR model.

Variables	Logistic Regression	FMLR(K = 2)	
	Parameter	component 1(0.672)	Component 2(0.328)
V_n^i	-	-0.1810*	0.1063*
ΔV_{nPL}^i	-0.40848*	-0.3903*	-0.2557*
D_n^i	.05490*	0.1895*	0.0158*
Y_n^i	.01345*	0.0109*	0.0105*
V_{nLead}^i	-.07111*	-0.0400	-0.0568*
ΔD_{nlg}^i	.01370*	0.0037	0.0113*
Constant	1.26281*	0.8417*	-1.7619*

Note: * means that the parameters are significant at 99% level.

TABLE 6: Mean values and standard deviations of related variables in each component.

Variables	Component 1		Component 2	
	Rejected Gaps (Standard Deviation)	Accepted Gaps (Standard Deviation)	Rejected Gaps (Standard Deviation)	Accepted Gaps (Standard Deviation)
V_n^i (m/s)	15.050 (3.196)	13.418 (3.107)	13.505 (2.852)	14.272 (3.466)
ΔV_{nPL}^i (m/s)	8.611 (3.825)	1.985 (2.766)	5.375 (3.698)	3.187 (2.908)
D_n^i (m)	10.068 (5.274)	33.14 (22.32)	17.468 (15.175)	27.09 (23.65)
V_{nLead}^i (m/s)	9.841 (7.114)	11.627 (6.671)	10.529 (6.420)	8.062 (8.100)
ΔD_{nlg}^i (m)	43.42 (47.04)	44.83 (42.99)	33.05 (35.06)	26.79 (35.46)
Merge Location(m)	41.66 (57.87)		108.58 (64.19)	
Number of Rejected Gaps	1.05		3.19	

and 3.19 and the average merge location is 41.66m and 108.58m, which indicates that drivers in Component 2 tend to choose gaps further downstream and rejected more gaps than drivers in Component 1. The average rejected gap of Component 2 (17.468 m) is much bigger than Component 1 (10.068m) while the average accepted gap of Component 2 (27.09 m) is much smaller than Component 1 (33.14 m), indicating the inconsistency of gap acceptance theories. One can also find that the drivers in Component 2 increase their speeds during their merging process from 13.505m/s to 14.272 m/s, while drivers in Component 1 decrease their speed from 15.050 m/s to 13.418m/s, and in Component 2, the speed difference between the putative leading vehicle and the merging vehicle for accepted gaps is 3.187m/s, which is much bigger than Component 1, both of which indicate that drivers in Component 2 are more aggressive than Component 1. It is interesting to find that the standard deviations of the speeds for rejected gaps and accepted gaps in Component 1 are similar, which is not the case in Component 2. And one can also find that the standard deviation of rejected gaps for Component 2 is much bigger than that in Component 1. These findings indicate that the merging process of drivers in Component 2 is much more complicated than drivers in Component 1.

Figure 2 shows the relation between the gap size and location for the rejected and accepted gaps in the two components. One can find that the accepted gaps of drivers in Component 1 are almost all located in the beginning

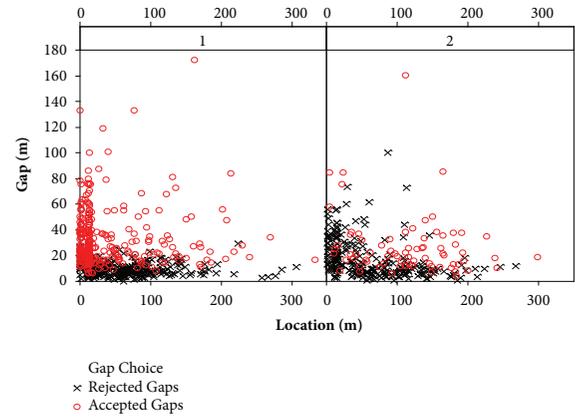


FIGURE 2: Relation between the gap size and location for the rejected and accepted gaps.

of the auxiliary lane while the accepted gaps of drivers in Component 2 are scattered along the lane. It is obvious that the rejected gaps of drivers in Component 2 are much larger than in Component 1 and are overlapped with the rejected gaps, while the overlapping area in Component 1 is much smaller.

Figure 3 shows the box plot of the reverse succession of offered gaps. The x-axis in Figure 3 is the reverse number of offered gaps before merging, in which 0 means the finally accepted gap and 1 means the last rejected gap before

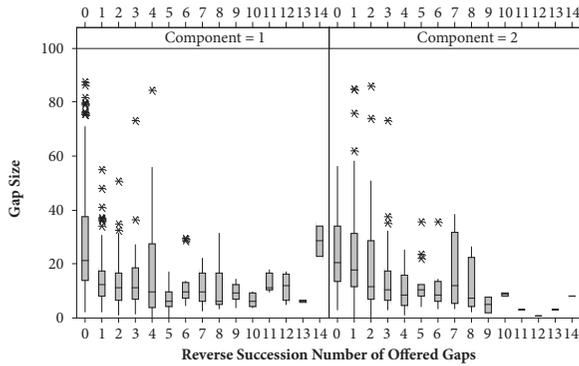


FIGURE 3: Box plot of the reverse succession of offered gaps.

TABLE 7: Comparison of estimated and observed values of logistic regression model.

		Estimated		Total
Observed	Reject	Accept		
Reject	484.0	68.0		552.0
Accept	92.0	281.0		373.0
Total	576.0	349.0		925.0

TABLE 8: Comparison of estimated and observed values of FMLR-2 model.

		Estimated		Total
Observed	Reject	Accept		
Reject	510.0	42.0		552.0
Accept	39.0	334.0		373.0
Total	547.0	378.0		925.0

merging. One can find that drivers in Component 2 might have several choices before merging, which indicates that drivers in Component 2 prefer to use the auxiliary lane to get further downstream.

Comparing the two components, drivers in Component 1 prefer larger gaps and lower speed difference, while drivers in Component 2 pay more attention to better surrounding traffic conditions and may sacrifice larger gaps to save travel time and get better traffic conditions. Thus, in this paper, Component 1 is named as Risk-Rejecting Drivers and Component 2 is named as Risk-Taking Drivers.

5.3. Accuracy of Developed Models. Tables 7 and 8 show the comparison of estimated and observed values of logistic regression model and 2-component mixture of logistic regression (FMLR-2) model. From these tables, the proposed model improves the predicting accuracy from 82.70% to 91.24%. It can be concluded that the proposed model has better predictive power than logistic regression model.

6. Conclusions

To incorporate the unobserved heterogeneity into merge model, the present study builds a FMLR model which uses BIC to determine the proper number of mixing components

and performs parameter estimation by using Latent GOLD 5.0.

Given U.S. Highway 101 data, the identified optimal model is a 2-component mixture of logistic regression model, which means the drivers can be divided into two components characterized by the driving behavior heterogeneity. One is the Risk-Rejecting Drivers whose drivers are consistent with previous studies and primarily merge in as soon as possible. Drivers in this component have a distinct preference for the larger gaps. The decrease of speed difference between merging vehicle and putative leading vehicle and a gap located further towards the end of the auxiliary lane also increase the probability of accepting the current gap. Contrast to Component 1, Component 2 is constituted with the drivers that are much less sensitive to the gap size and have more emphasis on surrounding traffic conditions such as the speed of front vehicle in the auxiliary lane and space gap between the merging vehicle and its leading vehicles in the auxiliary lane. These drivers are using the auxiliary lane to get to the further downstream or less congested area of the main lane. Thus they are called Risk-Taking Drivers.

In addition, the proposed model can produce more precise predicting accuracy than logistic regression model.

However, more empirical studies are needed to apply this method to datasets in other sites with different demographics, climate, and geometric parameters in order to fully assess the effect of the factors affecting merging behaviors as well as fully understand the strengths and weaknesses of the proposed model.

Data Availability

The NGISM data used to support the findings of this study have been deposited at the website: <https://catalog.data.gov/dataset/next-generation-simulation-ngsim-vehicle-trajectories>.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

- [1] L. Elefteriadou, R. P. Roess, and W. R. McShane, "Probabilistic nature of breakdown at freeway merge junctions," *Transportation Research Record*, vol. 1484, pp. 80–89, 1995.
- [2] H. Yi and T. E. Mulinazzi, "Urban freeway on-ramps: Invasive influences on main-line operations," *Transportation Research Record*, no. 2023, pp. 112–119, 2007.
- [3] S. P. Hoogendoorn and R. Hoogendoorn, "Generic calibration framework for joint estimation of car-following models by using microscopic data," *Transportation Research Record*, no. 2188, pp. 37–45, 2010.
- [4] J. Kim and H. S. Mahmassani, "Correlated parameters in driving behavior models," *Transportation Research Record*, vol. 2249, pp. 62–77, 2011.
- [5] S. Ossen and S. P. Hoogendoorn, "Heterogeneity in car-following behavior: theory and empirics," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 2, pp. 182–195, 2011.

- [6] I. Kim, T. Kim, and K. Sohn, "Identifying driver heterogeneity in car-following based on a random coefficient model," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 34–44, 2013.
- [7] Cambridge Systematics, *NGSIM US 101 Data Analysis: Summary Report*, Prepared for Federal Highway Administration, 2005.
- [8] Q. Yang and H. N. Koutsopoulos, "A microscopic traffic simulator for evaluation of dynamic traffic management systems," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 3, pp. 113–129, 1996.
- [9] K. I. Ahmed, *Modeling drivers' acceleration and lane changing behavior*, Massachusetts Institute of Technology, 1999.
- [10] G. Lee, *Modeling gap acceptance at freeway merges*, Massachusetts Institute of Technology, 2006.
- [11] T. Toledo, H. N. Koutsopoulos, and M. Ben-Akiva, "Integrated driving behavior modeling," *Transportation Research Part C: Emerging Technologies*, vol. 15, no. 2, pp. 96–112, 2007.
- [12] M. A. Ahammed, Y. Hassan, and T. A. Sayed, "Modeling driver behavior and safety on freeway merging areas," *Journal of Transportation Engineering*, vol. 134, no. 9, pp. 370–377, 2008.
- [13] T. Toledo, H. N. Koutsopoulos, and M. Ben-Akiva, "Estimation of an integrated driving behavior model," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 4, pp. 365–380, 2009.
- [14] F. Marczak, W. Daamen, and C. Buisson, "Merging behaviour: Empirical comparison between two sites and new theory development," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 530–546, 2013.
- [15] D. C. Gazis, R. Herman, and R. W. Rothery, "Nonlinear follow-the-leader models of traffic flow," *Operations Research*, vol. 9, no. 4, pp. 545–567, 1961.
- [16] A. J. Miller, *Nine estimators of gap-acceptance parameters*, Publication of: Traffic Flow and Transportation, 1971.
- [17] H. Mahmassani and Y. Sheffi, "Using gap sequences to estimate gap acceptance functions," *Transportation Research Part B: Methodological*, vol. 15, no. 3, pp. 143–148, 1981.
- [18] P. G. Gipps, "A model for the structure of lane-changing decisions," *Transportation Research Part B: Methodological*, vol. 20, no. 5, pp. 403–414, 1986.
- [19] P. Hidas, "Modelling lane changing and merging in microscopic traffic simulation," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 5-6, pp. 351–371, 2002.
- [20] J. Wang, "A simulation model for motorway merging behaviour," *Transportation and traffic theory*, vol. 16, pp. 281–301, 2005.
- [21] SiAS, *S-Paramics 2005 Reference Manual*, SIAS Ltd Edinburgh, 2005.
- [22] PTV-Vision, *VISSIM 5.30-05 user manual*, 2011.
- [23] W. Daamen, S. P. Hoogendoorn, and M. Looij, "Empirical analysis of merging behavior at freeway on-ramp," *Transportation Research Record: Journal of Transportation Research Board*, no. 2188, pp. 108–118, 2010.
- [24] T. D. CHU, *A Study on Merging Behavior at Urban Expressway Merging Sections*, Nagoya University, 2014.
- [25] H. Kita, "Effects of merging lane length on the merging behavior at expressway on-ramps," *Transportation and Traffic Theory*, pp. 37–51, 1993.
- [26] H. Kita, "A merging-giveway interaction model of cars in a merging section: a game theoretic analysis," *Transportation Research Part A: Policy and Practice*, vol. 33, no. 3-4, pp. 305–312, 1999.
- [27] J. Weng and Q. Meng, "Modeling speed-flow relationship and merging behavior in work zone merging areas," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 985–996, 2011.
- [28] A. Rao, *Modeling Anticipatory Driving Behavior [MS Thesis]*, Department of Civil and Environmental Engineering, MIT, 2006.
- [29] C. F. Choudhury, M. E. Ben-Akiva, T. Toledo, G. Lee, and A. Rao, "Modeling cooperative lane changing and forced merging behavior," in *Proceedings of the 86th Annual Meeting of the Transportation Research Board, C., Ed.*, Washington, DC, 2007.
- [30] L. Sun and J. Zhou, "Development of multiregime speed-density relationships by cluster analysis," *Transportation Research Record*, no. 1934, pp. 64–71, 2005.
- [31] Y. Pan and L. Sun, "Characterizing heterogeneity in vehicular traffic speed using two-step cluster analysis," *Journal of Southeast University (English Edition)*, vol. 28, no. 4, pp. 480–484, 2012.
- [32] S. Ossen and S. P. Hoogendoorn, "Car-following behavior analysis from microscopic trajectory data," *Transportation Research Record*, no. 1934, pp. 13–21, 2005.
- [33] S. Ossen, S. P. Hoogendoorn, and B. G. H. Gorte, "Interdriver differences in car-following a vehicle trajectory-based study," *Transportation Research Record*, vol. 1965, no. 1, pp. 121–129, 2006.
- [34] X. Ma and I. Andréasson, "Statistical analysis of driver behavior data in different regimes of the car-following stage," *Transportation Research Record*, no. 2018, pp. 87–96, 2007.
- [35] G. Li and L. Sun, "Characterizing Heterogeneity in Drivers Merging Maneuvers Using Two-Step Cluster Analysis," *Journal of Advanced Transportation*, 2018.
- [36] M. Keyvan-Ekbatani, V. L. Knoop, and W. Daamen, "Categorization of the lane change decision process on freeways," *Transportation Research Part C: Emerging Technologies*, vol. 69, pp. 515–526, 2016.
- [37] V. Alexiadis, J. Colyar, J. Halkias, R. Hranac, and G. McHale, "The next generation simulation program," *ITE Journal*, vol. 74, no. 8, pp. 22–26, 2004.
- [38] C. Thiemann, M. Treiber, and A. Kesting, "Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data," *Transportation Research Record*, vol. 2088, pp. 90–101, 2008.
- [39] V. Punzo, M. T. Borzacchiello, and B. Ciuffo, "On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1243–1262, 2011.
- [40] M. Montanino and V. Punzo, "Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns," *Transportation Research Part B: Methodological*, vol. 80, pp. 82–106, 2015.
- [41] B. Coifman and L. Li, "A critical evaluation of the Next Generation Simulation (NGSIM) vehicle trajectory dataset," *Transportation Research Part B: Methodological*, vol. 105, pp. 362–377, 2017.
- [42] Q. Wang, Z. Li, and L. Li, "Investigation of discretionary lane-change characteristics using next-generation simulation data sets," *Journal of Intelligent Transportation Systems*, vol. 18, pp. 246–253, 2014.
- [43] X. Wan, P. J. Jin, F. Yang, and B. Ran, "Merging preparation behavior of drivers: How they choose and approach their merge

- positions at a congested weaving area,” *Journal of Transportation Engineering*, vol. 142, no. 9, 2016.
- [44] X. Wan, P. J. Jin, H. Gu, X. Chen, and B. Ran, “Modeling Freeway Merging in a Weaving Section as a Sequential Decision-Making Process,” *Journal of Transportation Engineering, Part A: Systems*, vol. 143, Article ID 05017002, 2017.
- [45] G. McLachlan and D. Peel, *Finite mixture models, wiley series in probability and statistics*, John Wiley & Sons, New York, USA, 2000.
- [46] C. Fraley and A. E. Raftery, “Model-based clustering, discriminant analysis, and density estimation,” *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, 2002.
- [47] W. S. DeSarbo and W. L. Cron, “A maximum likelihood methodology for clusterwise linear regression,” *Journal of Classification*, vol. 5, no. 2, pp. 249–282, 1988.
- [48] M. Wedel and W. S. DeSarbo, “A mixture likelihood approach for generalized linear models,” *Journal of Classification*, vol. 12, no. 1, pp. 21–55, 1995.
- [49] M. Wedel and W. A. Kamakura, *Market segmentation: Conceptual and methodological foundations*, Springer Science & Business Media, 2012.
- [50] W. H. Greene and D. A. Hensher, “A latent class model for discrete choice analysis: contrasts with mixed logit,” *Transportation Research Part B: Methodological*, vol. 37, no. 8, pp. 681–698, 2003.
- [51] W. H. Greene, *Interpreting estimated parameters and measuring individual heterogeneity in random coefficient models*, 2004, Interpreting estimated parameters and measuring individual heterogeneity in random coefficient models.
- [52] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society Series: B*, vol. 39, no. 1, pp. 1–38, 1977.
- [53] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*, John Wiley & Sons, 2007.
- [54] J. K. Vermunt and J. Magidson, *Latent GOLD 5.0 upgrade manual*, Statistical Innovations Inc, Belmont, MA, 2013.
- [55] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, 1974.
- [56] G. Schwarz, “Estimating the dimension of a model,” *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

Research Article

Characterizing Critical Transition State for Network Fundamental Diagram

Rongrong Hong,¹ Chengchuan An,¹ Zhenbo Lu ,¹ Jingxin Xia ,¹ Qinghui Nie,² and Wenming Rao¹

¹Intelligent Transportation Systems Research Center, Southeast University, Nanjing, China

²College of Civil Science and Engineering, Yangzhou University, Yangzhou, China

Correspondence should be addressed to Zhenbo Lu; seuitslzb@126.com

Received 8 June 2018; Revised 10 September 2018; Accepted 24 October 2018; Published 18 November 2018

Guest Editor: Lele Zhang

Copyright © 2018 Rongrong Hong 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.

Macroscopic Fundamental Diagram (MFD) reveals the relationship between network accumulation and flow at the macroscopic level. The network traffic flow state analysis is a fundamental problem for the MFD-based applications. Theoretical and experimental investigations have provided insights into the dynamics and characters of traffic flow states. Although many empirical studies had been conducted in the field of MFD, few studies were dedicated to investigate the network traffic flow states with field data. This study aims to develop a data-driven method based on time series analysis of MFD state points to characterize critical transition state (CTS) of network traffic flow using field data. The proposed method was tested in a real network of Kunshan City, China. The test results showed that the CTS points can be well captured by the proposed method. The identified CTS points distinguished the traffic states between free-flow state and optimal accumulation state, and the optimal accumulation state was characterized. The day-to-day pattern of CTS points was investigated by the Gaussian Mixture Model-based clustering model. An extended application of real-time identification of CTS points was also discussed. The proposed method is helpful to understand the temporal evolution process of network traffic flow and provides potentials for developing more reliable network traffic flow management strategies, such as optimizing traffic signal plans and developing strategies for congestion tooling.

1. Introduction

Macroscopic Fundamental Diagram (MFD) reveals the relationship between network accumulation and flow at the macroscopic level, and can be readily obtained with traffic flow data from existing detectors (e.g., loop detectors). The core property of MFD is that the network flow is maximized when network traffic stays at an optimal accumulation state. This property can be used to develop effective applications such as traffic signal gating control [1–3], congestion pricing [4], route guidance [5] and traffic safety analysis [6]. To the successes of these applications, one of the important bases is to understand the characteristic of network traffic flow states in MFD, including the state transitions in the temporal evolution process.

To characterize the network traffic flow states with MFD, analytical methods were proposed in theoretical studies. Daganzo and Gayah [7] illustrated that bifurcations occurred

when the network accumulation came up to a certain value. In light of their work, a method based on the phase portraits of the dynamic Equations was developed to determine the boundaries of the stable and unstable regions of states in MFD [8]. Gayah and Daganzo [9] found that the clockwise hysteresis phenomenon during the congestion recovery period was caused by the instability of network traffic flow. Daganzo and Geroliminis [10] developed an analytical approximation method to derive the upper bound of the average flow conditional on average density. The derived network flow was interpreted as the theoretical capacity of network. Following their work, Geroliminis and Boyac [11] investigated the effects of different link lengths and signal timing on network capacity. Although these studies have revealed important properties of macroscopic network traffic flow states, the theoretical methods were not ready to apply in practice due to limitations such as the overestimation of network capacity [12, 13].

More practically, empirical studies using field data gained more attentions. In literature, most studies focused on the estimation of MFD with different sources of data, including loop detector data [14, 15], probe vehicle data [15–17] and connected vehicle data [18]. However, few studies dedicated to develop methods for the analysis of network traffic flow states, such as the network traffic flow state identification and state evolution pattern recognition. Saberi and Mahmassani [19] conducted a comparison analysis to investigate the capacity drop phenomena using loop detector data from three different freeway networks. The optimal accumulation states were typically identified by observation [20, 21]. A recent study by Ampountolas and Kouvelas [22] proposed a Kalman Filter based method to identify the maximum network flow according to the changes of slopes between consecutive points in MFD.

In contrast to the limited analysis of network traffic flow states, the analysis of link traffic flow states was extensively discussed using field data in the past decades. The equilibrium relationship among the traffic flow, density, and speed on links of urban roads and freeways was obtained by fitting the scatter state points in link traffic flow fundamental diagram [23–25]. Further, clustering methods were applied to the scatter state points to identify the traffic flow states. The methods include nested clustering [26], fuzzy *c*-means clustering [27, 28], online agglomerative clustering [29], K-means [30], multivariate clustering [31], and DBSCAN clustering [32]. These studies intended to answer the questions that how many traffic states could be described, and what the pattern of traffic flow is within each state. However, the temporal evolution patterns of traffic states are not specifically addressed in the clustering methods. Thus, in another perspective, analysis methods were developed to identify the transition state on the time series of the traffic flow states. The key procedure of these studies is to identify the breakpoints in a time continuous traffic flow state series. Banks [33] measured the significant changes of the slopes between consecutive flow-occupancy state points and identified the transition states between a free-flow state and a synchronized flow state. Vlahogianni et al. [34] used the Recurrence Plot (RP) to identify the point where a shift or abrupt change in the traffic volume series taken place. Kamarianakis et al. [35] used a smooth-transition regression to identify an asymmetric transition behaviour between free-flow and congestion state. Tang et al. [36] and Yan et al. [37] reconstructed the time series of traffic flow, occupancy, and speed as a network graph and identified different traffic flow states within the time series based on a complex network method. Based on the analysis of the time series of traffic flow states, Blandin et al. [38] proposed a modified LWR model to describe the temporal transition characteristic of traffic flow states. Despite the similarity of the problems between the link and network traffic flow state analysis, the characteristic of traffic flow states on network could be very different from that on a single link.

The problem addressed in this study is to identify and characterize the critical transition state (CTS) point in MFD using field data. The CTS point is defined as the point where significant changes occurs in a MFD time series, meaning that the network traffic flow state before and after a CTS

point is significantly different. First, the similarity of two subseries before and after a point in a MFD time series was measured. Second, the CTS points were identified based on the similarity measurement. Third, the day-to-day patterns were investigated based on the identified CTS points. The proposed method was designed to analyse network traffic flow states based on historical data. In addition, an extended application of real-time identification of CTS points was also discussed in the end. The proposed method is helpful to better understand the evolution process of network traffic flow and offers opportunities to develop more reliable and responsive network traffic flow management strategies. For example, with the identification of CTS point, it is helpful to determine when to use the peak control strategy to improve traffic flow conditions in real time. Besides, it also can be used to determine to the best period to implement congestion tooling policy before the urban traffic flow becomes too congested.

2. Methodology

2.1. Framework. The framework of the methodology involves three parts, as illustrated in Figure 1. The first part is to measure the similarity between two subseries around each point in a MFD time series. The two subseries are extracted by a sliding window structure, and the Dynamic Time Wrapping (DTW) distance [39] is used as the similarity measurement. The second part is to find the most significant DTW distance through a peak searching process to identify the CTS point. The Locally Weighted Regression (LOESS) [40] is initially used to smooth the DTW distance series extracted in the first part. Then, the peaks of the smoothed DTW distance series are determined based on a root-finding method [41]. The third part is to analyse the day-to-day patterns of the CTS points identified from MFD time series of different days. The patterns are recognized by a Gaussian Mixture Model (GMM) clustering method [42].

2.2. Measuring Pattern Similarity with Sliding Windows. Let s_t be a state point in a MFD time series for the t th interval in a day, and s_t is a two-dimensional vector defined as the network occupancy and network flow (occ_t, vol_t). The network occupancy and network flow are the average occupancy over each link and average volume weighted by link length in a network [14].

The two subseries before and after s_t are defined in two sliding windows with the length l , as $SB_t = (s_{t-l}, s_{t+1-l} \dots s_{t-1}) = (sb_1 \dots sb_l \dots sb_l)$ and $SA_t = (s_t, s_{t+1} \dots s_{t+l-1}) = (sa_1 \dots sa_j \dots sa_l)$, respectively. The similarity measurement between the two subseries is notated as $d(SB_t, SA_t)$ and attributed to s_t . SB_t and SA_t are normalized to ensure the two subseries are comparable. The Dynamic Time Wrapping (DTW) distance is used to calculate $d(SB_t, SA_t)$. The reason that DTW distance is adopted as the similarity measurement of the two subseries rather than Euclidean distance is that the DTW distance is able to compare the trends (e.g., increasing or decreasing trends) between two series. Let sb_i and sa_j be the i th and j th element of SB_t and SA_t , respectively.

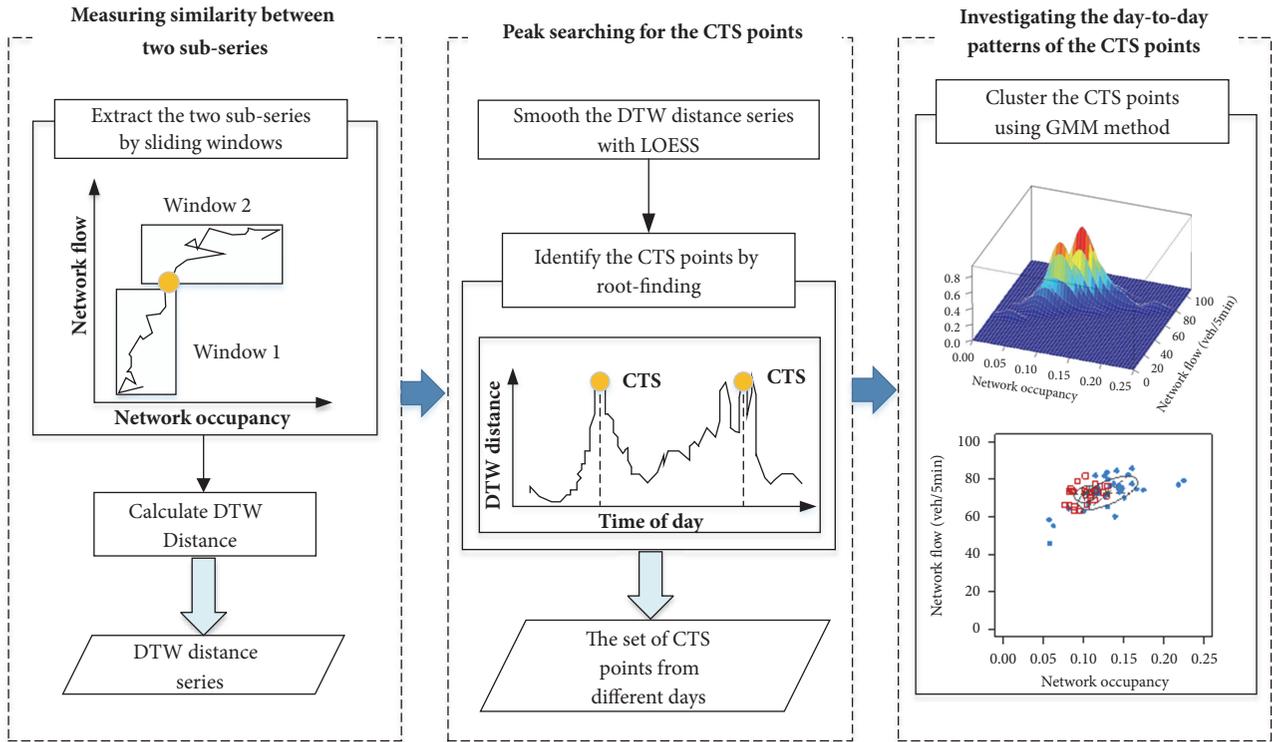


FIGURE 1: Framework of proposed method.

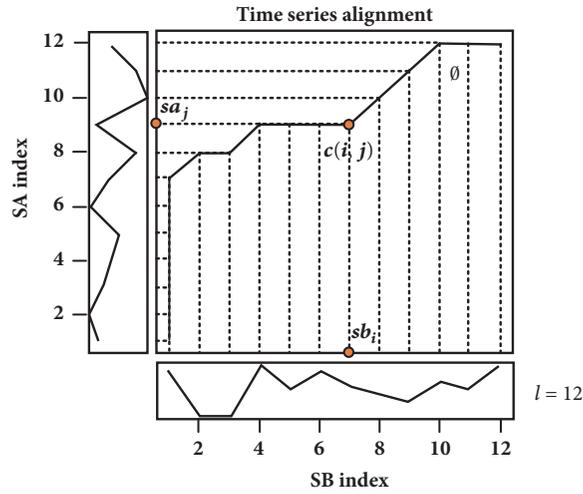


FIGURE 2: An optimal correspondence on warped evolution paths for s_t .

The schematic diagram for calculation of the DTW distance is illustrated in Figure 2. The DTW distance is calculated in two steps. First, calculate the local cost matrix. Each item of the matrix with index (i, j) represents a local cost $c(i, j)$ between sb_i and sa_j . Second, calculate the accumulated cost matrix or global cost matrix, as shown in (1). Third, find the shortest path with the global cost matrix $\beta(i, j)$ shown as the path θ in Figure 2. The global cost $\beta(l, l)$ is the DTW distance between the two series. The DTW distance can be measured for two series with different lengths. In this study,

the lengths of two series are the same as l . Thus, the cost matrix is $l \times l$. The $c(i, j)$ is calculated as the Euclidean distance between sb_i and sa_j . To effectively find the shortest path, dynamic programming algorithm is used.

$$\begin{aligned} \beta(i, j) &= c(i, j) \\ &+ \min \{ \beta(i-1, j-1), \beta(i-1, j), \beta(i, j-1) \} \end{aligned} \tag{1}$$

where $\beta(i, j)$ is the global cost from item (1,1) to item (i, j). By iteratively solving (1), the shortest path can be obtained.

2.3. Peak Searching for Critical Transition State Point. Based on the DTW distance, the CTS points could be identified as the points with significantly large DTW distances. Intuitively, a threshold-based method could be used. Thus, all points with DTW distances larger than a predetermined threshold would be regarded as the CTS points. However, the selection of the thresholds could be difficult and arbitrary because different networks may have distinct thresholds. Basically, the identification of the CTS points is an extreme value analysis problem over a series. Therefore, a more sophisticated algorithm was developed in this study. The algorithm was described as follows.

Step 1. Smooth the distance series with Locally Weighted Regression (LOESS) [40]. The smoothing process is initially conducted in order to filter the occasionally and locally fluctuated points in the DTW distance series. Let y_j be the smoothed value of the j th item in the DTW distance series which were divided into n equally spaced intervals.

Step 2. Determine the ranges that contain the CTS points. The identification of CTS is to find the extreme point (EP) of DTW series. To determine the range of each extreme point (EP), the approximate extreme point (AEP) is captured first. The AEP locates at a small scale that an extreme point was contained. Notate Δy_j and $\Delta^2 y_j$ as the first and second order difference of y_j . The monotonicity of $y(j)$ is determined by comparing the sign of $y_j, \Delta y_j, \Delta^2 y_j$ which were expressed in (2) and (3).

$$\begin{aligned} y_j &= f(j), \\ \Delta y_j &= f(j+1) - f(j), \end{aligned} \quad (2)$$

$$\begin{aligned} \Delta^2 y_j &= \Delta y_{j+1} - \Delta y_j \\ \Delta y_j^* &= \begin{cases} 1, & \Delta y_j > 0 \\ 0, & \Delta y_j < 0 \end{cases} \end{aligned} \quad (3)$$

If $(\Delta^2 y_j) \neq 0$, the monotonicity of $f(j+1)$ changed compared with $f(j)$, and there is an AEP around $(j+1)$ th interval. Based on the determined AEPs, the ranges that contain EPs are determined by the centres of neighbouring AEPs.

Step 3. Identify the CTS points using Brend's method. The identification for CTS (EP) point can be viewed as rooting-finding of the first derivative of y within each range. Brend's method [41] is used to find the extreme point y_c with local maximum distance within each range. The method is a root-finding algorithm combining the bisection method, the secant method, and inverse quadratic interpolation which has the advantage of fast-converging. Corresponding to y_c , the critical state point s_c is achieved.

2.4. Day-to-Day Pattern Recognition of the CTS Points. Due to the stochastic nature of urban traffic flow, it is assumed that the day-to-day distribution of the CTS points is subject to a Gaussian Mixture Distribution. In this study, a Gaussian Mixture Model (GMM) is used to cluster the CTS points extracted from different days. Compared to other clustering methods such as the K-means, one appealing feature of the GMM clustering method is that the probability of a state point belonging to a cluster is available. The GMM clustering method is described as follows.

In GMM, the data is assumed to be generated from a mixture of a finite number of Gaussian distributions. Each Gaussian distribution is regarded as a component of the GMM, and the components are linearly combined by mixing weights. In fact, the GMM clustering is equivalent to estimate the unknown mixing weights and parameters of each Gaussian distribution. In this study, in addition to the network flow and occupancy, the temporal feature (e.g., time index during a day) of the CTS points is also included to the clustering process. Let s_c represents the set of the identified CTS points from multiple days. s_{ci} which represents for CTS point is a vector formed by network flow-occupancy and the time of day for the i th element in s_c . $\theta_k(s_{ci} | \theta_k)$ is the Probability Density Function (PDF) of k th component. The GMM model is shown in

$$\begin{aligned} \theta(s_{ci} | \theta) &= \sum_{k=1}^G \tau_k \theta_k(s_{ci} | \theta_k), \\ &\left(s_{ci} \in R^3, k \in \{1, 2, \dots, G\}, \tau_k \in (0, 1), \sum_k \tau_k = 1 \right) \end{aligned} \quad (4)$$

where k presents the k th component in the GMM, G is the total number of the components, $\theta_k = (\mu_k, \Sigma_k)$, μ_k and Σ_k are the mean and covariance of the k th component, and τ_k is the mixing weight of the k th component.

As described in (5), the GMM-based clustering method can be viewed as a process to find the optimal parameters τ_k and θ_k to maximize the likelihood function of GMM.

$$L(\theta; s_c, Z) = \prod_{i=1}^N \prod_{k=1}^G [\tau_k \theta_k(s_{ci} | \theta_k)]^{I(Z_i=k)} \quad (5)$$

θ_k and τ_k were iteratively estimated using the Expectation-Maximization (EM) algorithm [42]. The E-step which is expressed in (6) aims at getting the posterior probability ($\gamma(s_{ci}, k)$) of s_{ci} belonging to component k with initial θ_k and τ_k . The M-step is to get the new estimate of θ_k and τ_k through maximizing $L(\theta; s_c, Z)$ with posterior probability achieved in E-step. The estimation of parameters is shown in (7)-(9).

$$\gamma(s_{ci}, k) = \frac{\tau_k \theta_k(s_{ci} | \theta_k)}{\sum_{k=1}^G \tau_k \theta_k(s_{ci} | \theta_k)} \quad (6)$$

$$\mu_k = \frac{\sum_{i=1}^N \gamma(s_{ci}, k) s_{ci}}{\sum_{i=1}^N \gamma(s_{ci}, k)} \quad (7)$$

The clustering of critical transition state points based on GMM

Input: $s_c, \widehat{\theta}_t, G$

Output: $\mu_k^*, \sum_k^*, \tau_k^*$

Process:

- (1) Giving the initial $\widehat{\theta}_t$ and G
- (2) $E \leftarrow \widehat{\theta}_{t+1} - \widehat{\theta}_t$
- (3) **When** $E > \varepsilon$ **do**
- (4) $\gamma(s_{ci}, k) \leftarrow F(\widehat{\theta}_t)$ # step 4 is E-step (Equation (6))
- (5) $\widehat{\theta}_{t+1} \leftarrow \operatorname{argmax}_{\theta}$ # step 5 is M-step (Equation (7)-(9))
- (6) replace $\widehat{\theta}_t$ with $\widehat{\theta}_{t+1}$
- (7) **end for**

ALGORITHM 1: The EM Algorithm to estimate parameters in GMM.

$$\sum_k = \frac{\sum_{i=1}^N \gamma(s_{ci}, k) (s_{ci} - \mu_k) (s_{ci} - \mu_k)^T}{\sum_{i=1}^N \gamma(s_{ci}, k)} \quad (8)$$

$$\tau_k = \frac{\sum_{i=1}^N \gamma(s_{ci}, k)}{N} \quad (9)$$

where N is the total number of all critical transition state points. Z_j is a latent variable that determines the component from which the observation originates. The EM Algorithm is shown in Algorithm 1, $\widehat{\theta}_t$ is the initial value for θ , and t_i is the i th iteration. E is the difference between two neighbouring iterations, and ε is a small real number, and θ^* is the optimal value. It is worth mentioning that the BIC value is used to avoid overfitting and determine the optimal model parameters and the number of clusters [43].

3. Implementation

3.1. Data Collection. A real network with 13 signalized intersections and 26 links in Kunshan City, China, was selected as the test site, as shown in Figure 3. All intersections operated under pretimed signal control with four different signal plans in a day. The four arterials in the network had significant traffic volume during the day time. Oversaturated and spillover conditions were frequently observed during morning and afternoon peaks. Microwave-based detectors were installed on links except the ones shown as the dash link in Figure 3. All detectors were located approximately at the middle of the links and collected traffic counts and occupancy in every 5 minutes. Two months of data collected from August 1, 2015, to September 30, 2015, were used in this study. In addition, for validation purpose, GPS probe vehicle data from 1526 taxi cabs during the same period were also collected to calculate the network average speed. The space-mean speed of the network was simply calculated as the ratio between the total distance travelled and the total time spent by the taxi cabs in the test site for every 5 minutes.

3.2. Identification of the CTS Points. Before calculating DTW distance for the two subseries in the sliding windows, it is important to select a reasonable window length since the two subseries are expected to reflect the traffic evolution

patterns. However, due to the different OD patterns and signal operations in different networks, the window length is more likely network-dependent. In general, the window length should be shorter than the peak duration. Besides, the window length should not be too short to filter random fluctuations in data series and not too long to reflect traffic dynamics. Basically, determining a desirable length is a trade-off between reliability and accuracy.

In our case, the morning peak is from 6:00 to 9:00, and the afternoon peak is from 16:30 to 19:30. Therefore, the window length should be shorter than 3 hours. Based on that, the impacts of different window lengths on the measurement of the DTW distance were investigated. An index of Average Change Rate (ACR) was used to describe the fluctuation in the DTW distance series. The ACR is calculated as

$$ACR = \frac{1}{n} \sum_{i=1}^n \frac{d_{i+1} - d_i}{d_i} \quad (10)$$

where d_i is the normalized DTW distance associated with the i th state point in a MFD time series and n is the number of state points.

Figure 4(a) shows the measurement of DTW distance calculated from the MFD time series on August 10, 2015, with window length ranging from 15 minutes to 3 hours. It is observed that the curves of the DTW distance become smoother with the increasing of window length. The curves with window lengths of 1 hour and 1.5 hours are observed as the relatively balanced cases where the trends are clear and informative short-term changes are retained. In addition, the ACR was calculated using data in the two months, as shown in Figure 4(b). Each point in Figure 4(b) represents the ACR of the DTW distance series of a day with a particular window length. When the window length is lower than 1 hour, the ACR decrease significantly from about 0.11 to 0.01. A turning point can be clearly spotted at the window length of 1 hour. Based on the investigation, the window length was selected as 1 hour for our network.

With the measurement of DTW distance, a peak searching method was applied. Figure 5 illustrates the identified peaks on the DTW distance curve of August 11, 2015. The red and blue marks represent the local maximum and minimum

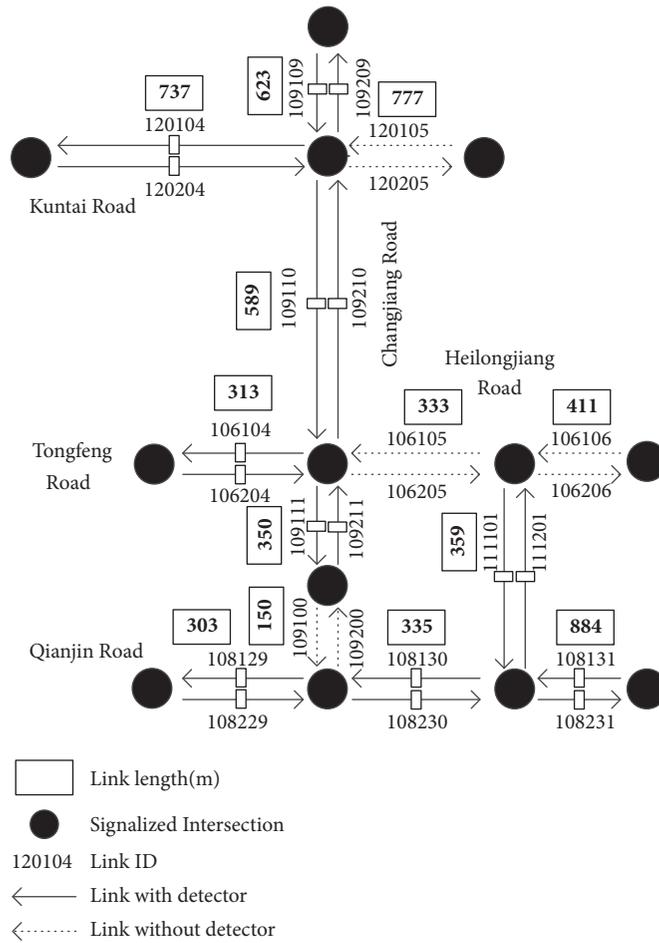


FIGURE 3: Layout of the test network and detectors in Kunshan City, China.

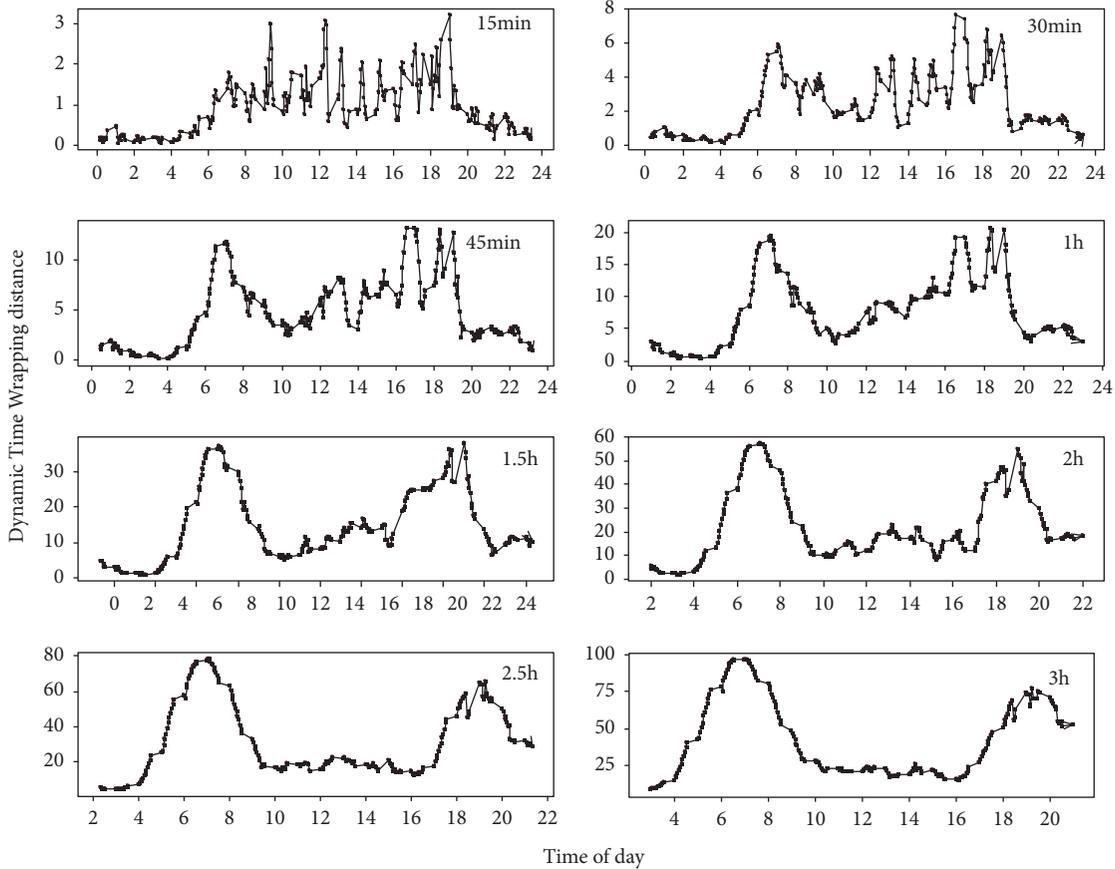
extreme points extracted by the proposed method. The local maximum extreme points are regarded as the CTS points in this study. As shown in Figure 5, the CTS points (see the points located at the high peaks of the curves) appear during the morning and afternoon peaks. Although the most critical transition state points are successfully identified, suspicious CTS points are also found at the local peaks on the curves. Some suspicious points (shown in the blue circle in Figure 5) are obviously invalid since the corresponding DTW distances are so small. These points could be readily excluded by a lower bound threshold. In this study, any state points with DTW distance less than 15 were considered as invalid points.

3.3. Validation of the CTS Points. To validate the identified CTS points, the locations of the CTS points at the MFD time series, the correlation between the CTS points, and network space-mean speed were investigated. For ease of illustration, the morning and afternoon peak were separately considered. The identified CTS points on August 10, Monday, 2015, and August 12, Wednesday, 2015, are illustrated in Figure 6. Figures 6(a), 6(d), and 6(g) show the CTS points during the morning peaks of the three days, and Figures 6(b), 6(e), and

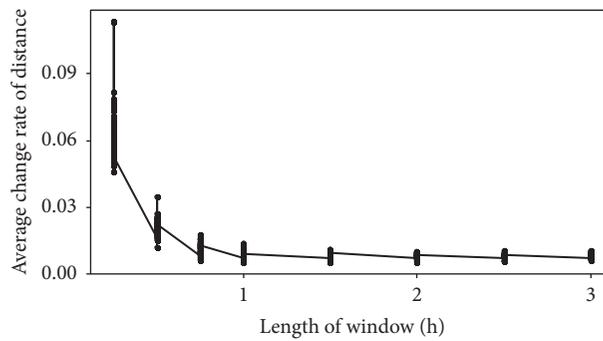
6(h) show the CTS points during the afternoon peaks of the three days. Figures 6(c), 6(f), and 6(i) illustrate the correlation between the CTS points and network space-mean speed for the three days. The red markers represent the identified CTS points. The arrows indicate the time sequence within the MFD time series.

For the locations at the MFD time series, the identified CTS points approximately locate at the boundaries of the two distinguished traffic states; see Figures 6(a), 6(d), 6(g), 6(b), 6(e), and 6(h). The state points on the left side of the CTS points show a positive correlation between network flow and occupancy, while the correlation of state points on the right side is vague and the slopes between the consecutive state points are much more fluctuated than the left side. A flat shape of the right side state points can be observed, and the network flow fluctuates within a small range from 75 to 90 veh/5min. Based on this, it is reasonable to infer that the left and right side of the identified CTS points are corresponding to the free-flow and optimal accumulation state, respectively. In other words, the identified CTS points well captured the shifts of network traffic states between the free-flow and optimal accumulation state.

By incorporating the network speed, in all these three days, the CTS points are identified at the times when the



(a)



(b)

FIGURE 4: Impacts of different window lengths on the measurement of DTW distance. (a) DTW distance curves with different lengths of window on August 10, Tuesday, 2015. (b) ACR calculation with different window lengths during August 1, 2015, and September 30, 2015.

network congestion just appeared and nearly recovered; see Figures 6(c), 6(f), and 6(i). Moreover, we can find that only the CTS points during the congestion formation stage are identified in the morning peaks of the three days. The reason may be that the traffic flow decreases slowly during the congestion recovery stage at the end of the morning peaks, and the traffic state after the morning peaks does not experience a significant change. According to the paired CTS points in afternoon peaks, the optimal accumulation state can be approximately estimated as (0.12, 73veh/5min) by averaging over the three days.

3.4. Day-to-Day Pattern Analysis. The empirical distribution of the identified CTS points based on two-month weekday data is visualized in as shown Figure 7. A clear bimodality can be observed in the distribution. In fact, the bimodality is caused by the different distributions of the CTS points in the morning and afternoon peaks.

Considering the most congested afternoon peaks in our case, day-to-day patterns of the CTS points from the afternoon peaks in 38 weekdays were investigated with the GMM clustering. Two distinct clusters are found, as shown in Figure 8. The cluster centre for Cluster 1 (in blue) is (17:18,

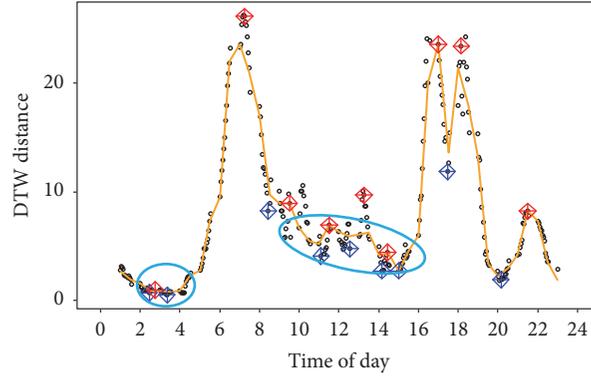


FIGURE 5: The DTW distance on August 11, Tuesday, 2015.

0.13, 72.8 (veh/5min)), and for Cluster 2 (in red) is (18:10, 0.10, 72.2(veh/5min)). 51.6% and 48.4% of the total identified CTS points are in Cluster 1 and Cluster 2, respectively. Figures 8(a) and 8(b) illustrate the distribution of flow-occupancy and time in the two clusters, respectively. According to the time distribution, Cluster 1 could be the critical transition state indicating a transition from the free-flow state to optimal accumulation state, while Cluster 2 is the critical transition state from the optimal accumulation state to free-flow state. In terms of the flow-occupancy distribution, it is found that the CTS points in Cluster 1 are much more scattered than those in Cluster 2. This indicates that the transition from the optimal accumulation state to free-flow state is much more predictable, and the transition from the free-flow to optimal accumulation state has higher uncertainty.

3.5. Extended Application of Real-Time CTS Point Identification. In addition to the analysis of the CTS points using historical data, an extended application of real-time CTS point identification is also discussed. Based on the proposed method, the MFD state points can be labelled as three classes: the CTS points of Cluster 1, the CTS points of Cluster 2, and non-CTS points. Thus, a probabilistic discriminant analysis [43] was conducted using the labelled dataset. The discriminant analysis, also known as supervised classification, assumes that the observations in each class are generated by a distribution specific to that class. Let θ_k be the probability distribution of the k th class which is Gaussian distribution in this study, τ_k be the proportion of observations of class k in the population, and s_t be a real-time observation of MFD state points. The posterior probability that s_t belongs to the k th class is expressed in

$$P_r(s_t \in \text{class } k) = \frac{\tau_k \theta_k(s_t; \mu_k, \Sigma_k)(s_t)}{\sum_{k=1}^G \tau_k \theta_k(s_t; \mu_k, \Sigma_k)(s_t)}, \quad (11)$$

$$k \in 1, 2, 3$$

In this study, the labelled MFD state points in the afternoon peaks of 45 days are used as the training set, and another 15 days as the test set. Each MFD state point was

identified as an observation of the class with the highest probability that belongs to. According to the test results, the error of the discriminant method is about 8.1%.

The discrimination results for three typical weekdays are illustrated in Figure 9. The black, red, and blue dots in Figure 9 represent the non-CTS points, the CTS points of Cluster 1, and the CTS point of Cluster 2, respectively. The probabilistic discriminant analysis performed well in identifying the CTS points. All the identified CTS points are located at the boundaries between the free-flow and optimal accumulation state, as shown in Figures 9(a), 9(c), and 9(e), and illustrate a strong correlation to the changes of network speed during the formation and recovery period of the network congestion.

4. Conclusions

This study proposed a data-driven method to identify and characterize the critical transition state in MFD using field data. The identification of the critical transition state is based on the analysis of the MFD time series. The DTW distance is used to measure the similarity of subseries before and after each state point in a MFD time series, and a peak searching method is developed to identify the critical transition state points based on the DTW distance series. The GMM clustering is used to characterize the day-to-day pattern of the identified critical transition state points. The probabilistic discriminant analysis is performed to identify the CTS points.

The identification and validation of CTS points indicated that the proposed method well identified the network transition states. By analysis, the identified CTS points well captured the shifts of network traffic states between the free-flow and optimal accumulation state. The optimal accumulation state can be approximately estimated as (0.12, 73veh/5min) by averaging over the three days. A clear bimodality can be observed by exploring the probability density of identified CTS points, and the bimodality is caused by different distributions of the CTS points in the morning and afternoon peaks. Moreover, the CTS points were classified as two categories. One cluster represents the transition from free-flow state to optimal accumulation state, and the other represents the transition from optimal accumulation state

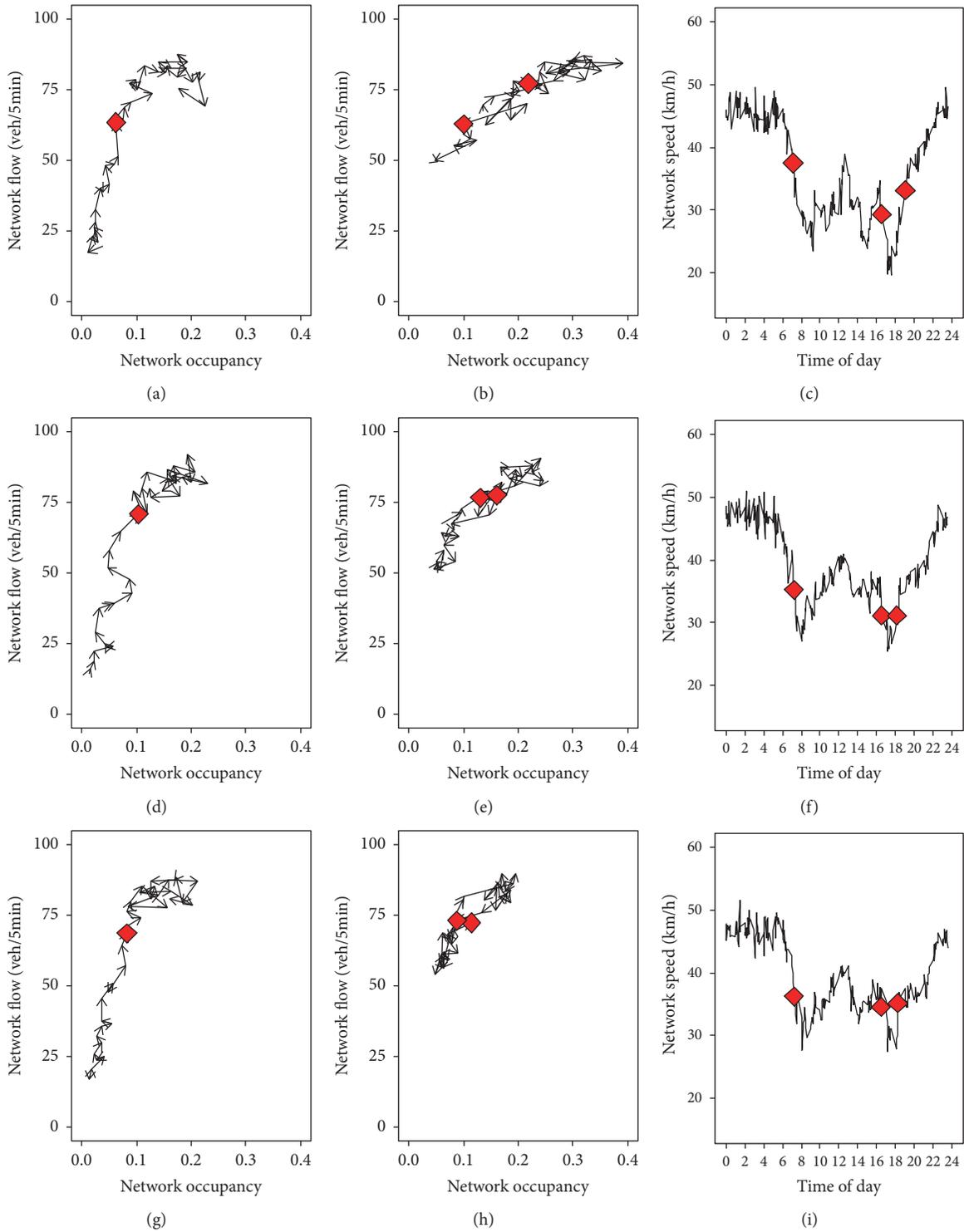


FIGURE 6: The validation of the identified CTS points on August 10, Monday, 2015, and August 12, Wednesday, 2015. (a), (d), and (g) The CTS points during morning peaks of the three days. (b), (e), and (h) The CTS points during afternoon peaks of the three days. (c), (f), and (i) Correlation between the CTS points and network speed.

to free-flow state. Besides, the probabilistic discriminant analysis performed well in identifying the CTS points. All the identified CTS points were located at the boundaries between the free-flow and optimal accumulation state. The proposed

method is helpful to better understand the evolution process of network traffic flow and offers opportunities to develop more reliable and responsive network traffic flow management strategies.

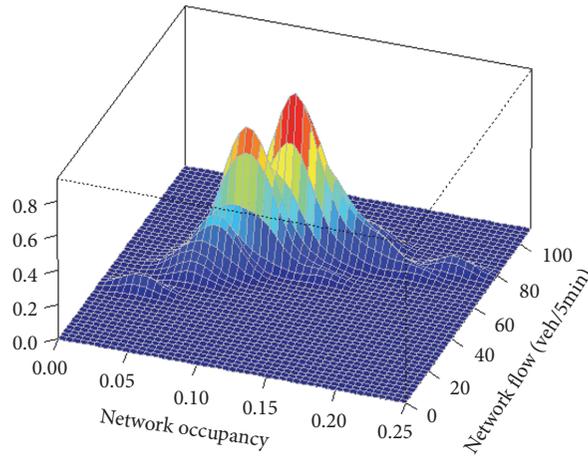


FIGURE 7: Visualization of the distribution of the CTS points.

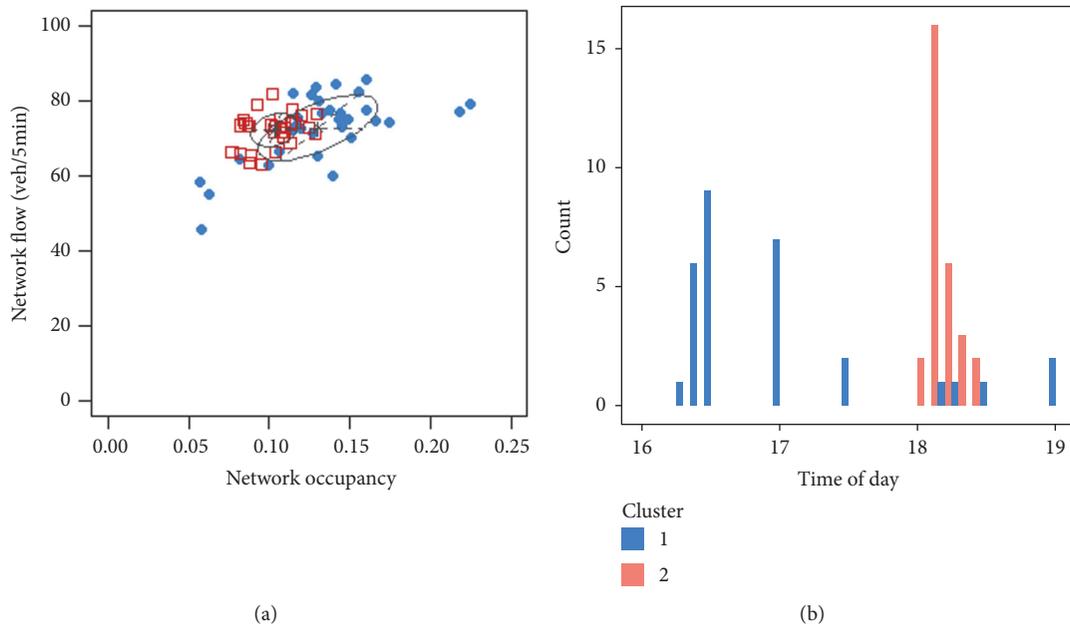


FIGURE 8: Day-to-day pattern of the CTS points from the afternoon peaks of 38 weekdays. (a) Flow-occupancy distribution in two clusters. (b) Time distribution in two clusters.

In future, the impact factors of critical transition states in actual network will be investigated through comparative analysis under certain conditions. Besides, it is encouraged to test the proposed method with different sources of data (e.g., probe vehicle or vehicle license plate reidentification data). Also, more field validation works are needed to further interpret the critical transition states considering real traffic behaviours.

Data Availability

Partial of the detector data used to support the findings of this study are included within the supplementary information file.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work was funded by the National Natural Science Foundation of China (no. 71871055) and the Key Research Program of Jiangsu Province Science and Technology Department (no. BE2017027).

Supplementary Materials

The supplementary material file is partial of the detector data used to derive the MFD for test site. The data are

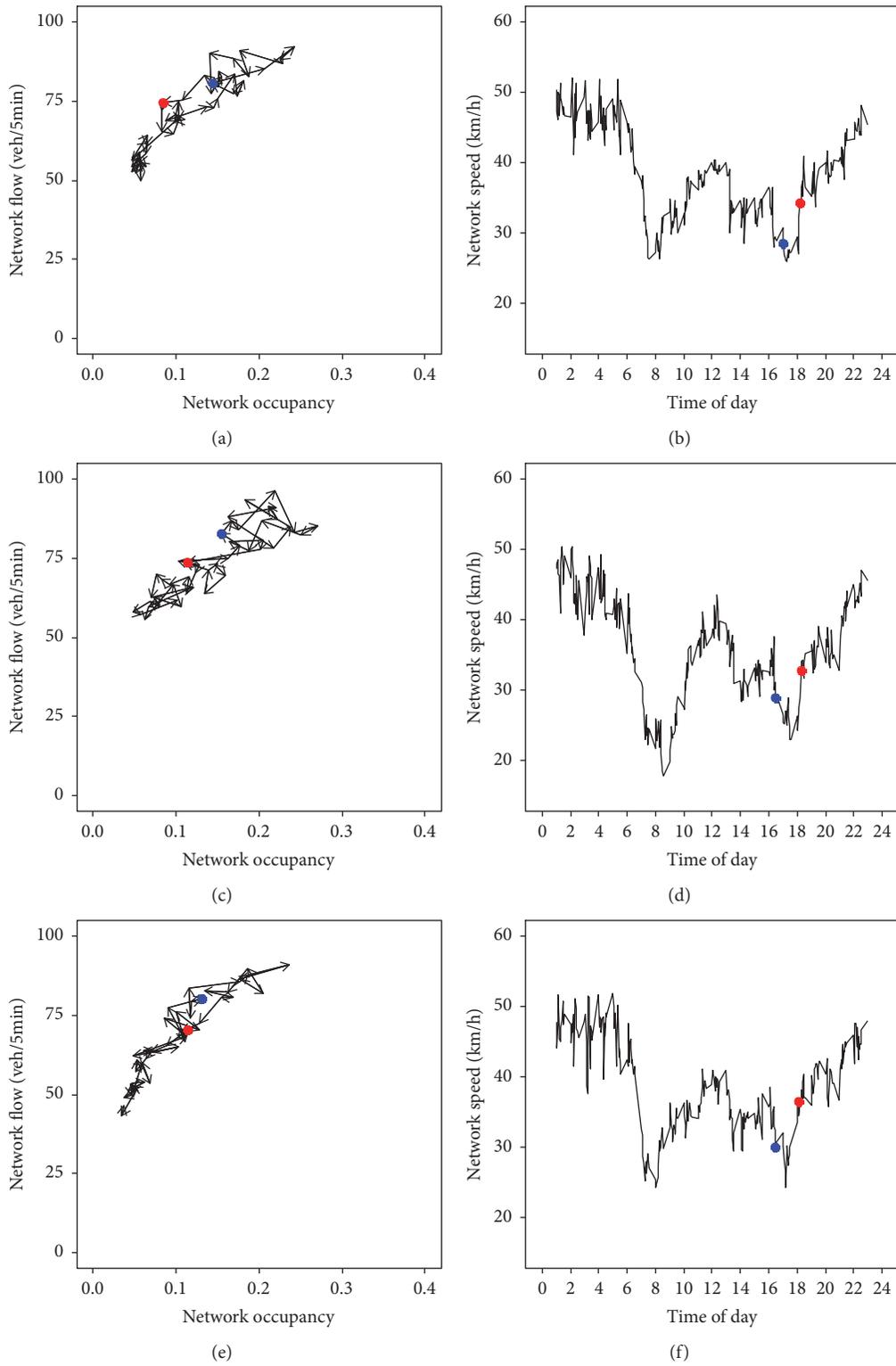


FIGURE 9: The discrimination results on three typical weekdays. (a), (c), and (e) The identified CTS points in the network flow-occupancy series during afternoon peak of the three days. (b), (d), and (f) The identified CTS points in the network speed series of the three days.

valid, and they were formed by STATIONID (the ID number of specific link), CALENDAR_KEY (the date of traffic flow data for specific link), FIVEMINX (the time of traffic flow data for specific date), OCCUPANCY (the occupancy/5min for specific link), and VOLUME (the flow/5min for specific link) from August 1, 2015, to September 30, 2015. (*Supplementary Materials*)

References

- [1] M. Keyvan-Ekbatani, M. Papageorgiou, and I. Papamichail, "Perimeter Traffic Control via Remote Feedback Gating," *Procedia - Social and Behavioral Sciences*, vol. 111, pp. 645–653, 2014.
- [2] Y. M. Du, J. P. Wu, and Y. Jia H, "MFD-Based Regional Traffic Volume Dynamic Control," *Journal of Transportation Systems Engineering Information Technology*, vol. 14, no. 3, pp. 162–167, 2014.
- [3] A. Csikós, T. Tettamanti, and I. Varga, "Macroscopic modeling and control of emission in urban road traffic networks," *Transport*, vol. 30, no. 2, pp. 152–161, 2015.
- [4] M. D. Simoni, A. J. Pel, R. A. Waraich, and S. P. Hoogendoorn, "Marginal cost congestion pricing based on the network fundamental diagram," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 221–238, 2015.
- [5] M. Hajiahmadi, V. L. Knoop, B. De Schutter, and H. Hellendoorn, "Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram," in *Proceedings of the 2013 16th International IEEE Conference on Intelligent Transportation Systems: Intelligent Transportation Systems for All Modes, ITSC 2013*, pp. 1022–1028, Netherlands, October 2013.
- [6] R. Alsalhi and V. Dixit, "The relationship between traffic safety and macroscopic fundamental diagram (MFD)," *Australasian Transport Research Forum (ATRF)*, vol. 37, 2015.
- [7] C. F. Daganzo, V. V. Gayah, and E. J. Gonzales, "Macroscopic relations of urban traffic variables: bifurcations, multivaluedness and instability," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 278–288, 2011.
- [8] J. Haddad and N. Geroliminis, "On the stability of traffic perimeter control in two-region urban cities," *Transportation Research Part B: Methodological*, vol. 46, no. 9, pp. 1159–1176, 2012.
- [9] V. V. Gayah and C. F. Daganzo, "Clockwise hysteresis loops in the Macroscopic Fundamental Diagram: An effect of network instability," *Transportation Research Part B: Methodological*, vol. 45, no. 4, pp. 643–655, 2011.
- [10] C. F. Daganzo and N. Geroliminis, "An analytical approximation for the macroscopic fundamental diagram of urban traffic," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 771–781, 2008.
- [11] N. Geroliminis and B. Boyaci, "The effect of variability of urban systems characteristics in the network capacity," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1607–1623, 2012.
- [12] E. J. Gonzales, C. Chavis, Y. Li, and C. F. Daganzo, *Multimodal transport modeling for nairobi, kenya: insights and recommendations with an evidence-based model*, insights and recommendations with an evidence-based model, kenya, 2009.
- [13] A. Mazloumian, N. Geroliminis, and D. Helbing, "The spatial variability of vehicle densities as determinant of urban network capacity," *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, vol. 368, no. 1928, pp. 4627–4647, 2010.
- [14] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.
- [15] T. Courbon and L. Leclercq, "Cross-comparison of macroscopic fundamental diagram estimation methods," in *Proceedings of the 14th Meeting of the Euro Working Group on Transp. - In Quest for Adv. Models, Tools, and Methods for Transp. and Logist., EWGT, 26th Mini-EURO Conf. - Intelligent Decis. Making in Transp. and Logist., MEC, 1st Eur. Sci. Conf. on Air Transp. - RH*, pp. 417–426, Poland, September 2011.
- [16] V. Gayah and V. Dixit, "Using mobile probe data and the macroscopic fundamental diagram to estimate network densities," *Transportation Research Record*, no. 2390, pp. 76–86, 2013.
- [17] J. Du, H. Rakha, and V. V. Gayah, "Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions," *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 136–149, 2016.
- [18] Z. Xu, P. J. Jin, and X. Jiang, "Macroscopic Fundamental Diagram (MFD) Based on Vehicle-to-Infrastructure (V2I) Connected Vehicle Data , Transportation Research Board 94th Annual Meeting," *Macroscopic Fundamental Diagram (MFD) Based on Vehicle-to-Infrastructure (V2I) Connected Vehicle Data , Transportation Research Board 94th Annual Meeting*, 2015.
- [19] M. Saberi and H. Mahmassani, "Hysteresis and capacity drop phenomena in freeway networks: Empirical characterization and interpretation," *Transportation Research Record*, no. 2391, pp. 44–55, 2013.
- [20] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, "Exploiting the fundamental diagram of urban networks for feedback-based gating," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1393–1403, 2012.
- [21] K. Aboudolas and N. Geroliminis, "Perimeter and boundary flow control in multi-reservoir heterogeneous networks," *Transportation Research Part B: Methodological*, vol. 55, pp. 265–281, 2013.
- [22] K. Ampountolas and A. Kouvelas, "Real-Time Estimation of Critical Values of the Macroscopic Fundamental Diagram for Maximum Network Throughput," *Transportation Research Board 94th Annual Meeting*, 2015.
- [23] B. Ponnu and B. Coifman, "When adjacent lane dependencies dominate the uncongested regime of the fundamental relationship," *Transportation Research Part B: Methodological*, vol. 104, pp. 602–615, 2017.
- [24] X. Qu, J. Zhang, and S. Wang, "On the stochastic fundamental diagram for freeway traffic: Model development, analytical properties, validation, and extensive applications," *Transportation Research Part B: Methodological*, vol. 104, pp. 256–271, 2017.
- [25] X. Qu, S. Wang, and J. Zhang, "On the fundamental diagram for freeway traffic: a novel calibration approach for single-regime models," *Transportation Research Part B: Methodological*, vol. 73, pp. 91–102, 2015.
- [26] J. Xia and M. Chen, "A nested clustering technique for freeway operating condition classification," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 6, pp. 430–437, 2007.
- [27] H.-p. Lu, Z.-y. Sun, and W.-c. Qu, "Big data-driven based real-time traffic flow state identification and prediction," *Discrete Dynamics in Nature and Society*, Art. ID 284906, 11 pages, 2015.

- [28] M. Azimi and Y. Zhang, "Categorizing Freeway Flow Conditions by Using Clustering Methods," *Transportation Research Record*, vol. 2173, no. 1, pp. 105–114, 2018.
- [29] J. Xia, W. Huang, and J. Guo, "A clustering approach to online freeway traffic state identification using ITS data," *KSCE Journal of Civil Engineering*, vol. 16, no. 3, pp. 426–432, 2012.
- [30] J. Tang, F. Liu, W. Zhang, S. Zhang, and Y. Wang, "Exploring dynamic property of traffic flow time series in multi-states based on complex networks: Phase space reconstruction versus visibility graph," *Physica A: Statistical Mechanics and its Applications*, vol. 450, pp. 635–648, 2016.
- [31] H. B. Celikoglu and M. A. Silgu, "Extension of traffic flow pattern dynamic classification by a macroscopic model using multivariate clustering," *Transportation Science*, vol. 50, no. 3, pp. 966–981, 2016.
- [32] S. Oh, Y.-J. Byon, and H. Yeo, "Impact of Traffic State Transition and Oscillation on Highway Performance with Section-Based Approach," in *Proceedings of the 18th IEEE International Conference on Intelligent Transportation Systems, ITSC 2015*, pp. 2141–2146, Spain, September 2015.
- [33] J. H. Banks, "Investigation of some characteristics of congested flow," *Transportation Research Record*, no. 1678, pp. 128–134, 1999.
- [34] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 5, pp. 351–367, 2006.
- [35] Y. Kamarianakis, H. Oliver Gao, and P. Prastacos, "Characterizing regimes in daily cycles of urban traffic using smooth-transition regressions," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 821–840, 2010.
- [36] J. Tang, Y. Wang, and F. Liu, "Characterizing traffic time series based on complex network theory," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 18, pp. 4192–4201, 2013.
- [37] Y. Yan, S. Zhang, J. Tang, and X. Wang, "Understanding characteristics in multivariate traffic flow time series from complex network structure," *Physica A: Statistical Mechanics and its Applications*, vol. 477, pp. 149–160, 2017.
- [38] S. Blandin, J. Argote, A. M. Bayen, and D. B. Work, "Phase transition model of non-stationary traffic flow: Definition, properties and solution method," *Transportation Research Part B: Methodological*, vol. 52, pp. 31–55, 2013.
- [39] D. F. Silva and G. E. Batista, "Speeding up all-pairwise dynamic time warping matrix calculation," in *Proceedings of the 2016 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics*, pp. 837–845, 2016.
- [40] O. B. Linton, "Local Regression Models," in *Microeconometrics*, Palgrave Macmillan, London, UK, 2010.
- [41] R. P. Brent, *Algorithms for Minimization without Derivatives*, Prentice Hall, Englewood Cliffs, NJ, USA, 2013.
- [42] N. M. Nasrabadi, "Pattern Recognition and Machine Learning," *Journal of Electronic Imaging*, vol. 16, no. 4, p. 049901, 2007.
- [43] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, 2002.

Research Article

Empirical Approximation for the Stochastic Fundamental Diagram of Traffic Flow on Signalized Intersection

Nan Zhang , Xiaoguang Yang , and Wangjing Ma 

Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Nan Zhang; zhangn_2007@163.com and Xiaoguang Yang; yangxg@tongji.edu.cn

Received 17 July 2018; Accepted 30 September 2018; Published 23 October 2018

Guest Editor: Lele Zhang

Copyright © 2018 Nan 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.

The wide scattering nature of the fundamental diagram (FD) with observed flow-density data may be associated with the dynamical traffic flow process, especially on signalized intersection. To describe the uncertainty of FD, in this work we established stochastic fundamental diagram (SFD) which is defined by the distributions of shockwave speed. Our approach is based on a two-level stochastic process of the traffic flow system in terms of the dynamics of traffic density and state mode associated with signal phases which is named switching linear dynamical systems (SLDS). Then, variational Bayesian learning method is adopted to compute the distributions of SFD parameter to approximate the experimental distributions of shockwave calculated by the observed flow-density data. Given traffic flow data from the NGSIM program, the verification result demonstrated that the SFD can be more helpful to capture the main features of the observed widely scattering of the flow-density data compared with FD. With the shockwave speed sampled from the SFD, the SLDS could describe the dynamic characteristics of traffic flow and be applied to the maximum likelihood estimation of traffic density or flow rate. Because it is simple and automatically calculated, the SFD provides an alternative description for fundamental diagram and its uncertainty in the traffic flow.

1. Introduction

The model-based traffic simulation and traffic state estimation require stochastic models to describe the dynamical phenomena of traffic flow system when the traffic flow management is applied on the freeways and signalized arterials. Compared to the randomness of driver behavior in microscopic simulation model, there are several methods to introduce stochastic elements to the macroscopic simulation model including the boundary conditions [1], the flow-density relationship or the fundamental diagram (FD) [2], and the traffic flow evolution process [3]. For the freeway condition, the scattered FD with observed flow-density data may be associated with the dynamical traffic flow system and the uncertainty of FD can be mostly expressed in terms of the variance of FD parameters [2, 4, 5]. Then, the problem is how to numerically determine the variance of FD parameters which can be derived from macroscopic simulation model [4, 5] or stochastic differential equation [2]. Considering the traffic flow of signalized intersection, widely scattered FD is more complicated than the freeway condition which may be

affected by not only the traffic flow system itself but also the traffic signals [6].

This research follows the abovementioned principle to explore the stochastic fundamental diagram (SFD) of traffic flow on signalized intersection. First, whether applied to the freeways or the signalized intersections, the common features of SFD are that the flow-density data is divided into two different regions: lower vehicle density and higher vehicle density, which correspond to “free” and “congested” flow. Once either of the density or the flow rate is determined, the other can be obtained by the “forward” and “backward” shockwave speed in each two different regions of “free” and “congested” flow. So, in our research, the SFD can be defined by the empirical distributions of traffic flow shockwave speed which is simple enough to be obtained and applied. Secondly, the macroscopic traffic simulation model, such as the cell transmission model (CTM) [7, 8], can succeed in describing traffic evolution under many different traffic conditions with the traffic flow density combined with FD. So, the traffic model can be employed to derive the SFD that should not be subjected to a specific traffic scenario or some certain

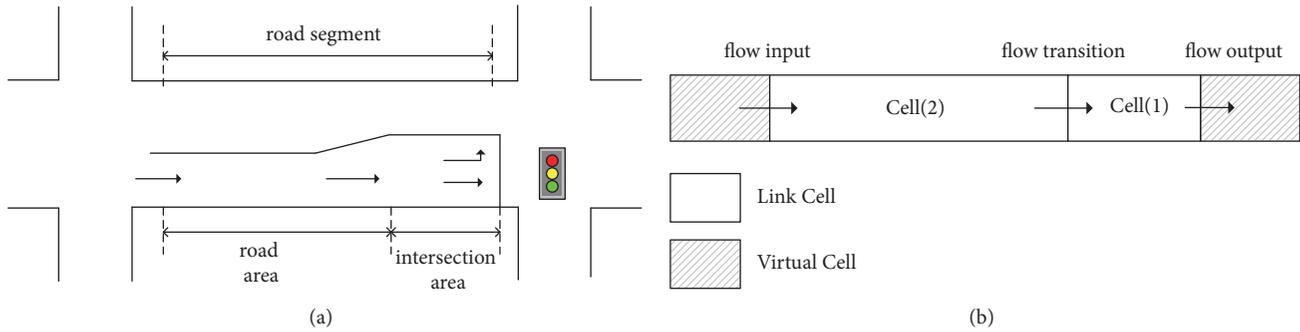


FIGURE 1: Cell representation of the road segment: (a) the signalized road segment, (b) the link cells of the road segment with virtual cells.

observed data. Alternatively, the SFD may be used as an input together with the boundary conditions to determine the temporal evolution of traffic flow system. Furthermore, in our research, traffic signal influence must be considered in traffic flow evolution process. For the freeway condition, complex traffic flow evolution process can be regarded as jump Markov linear systems in which the traffic state transitions between different modes with constant FD [9, 10]. Further, a stochastic cell transmission model (SCTM) has been proposed with stochastic demand and supply introduced in a switching-mode model [3]. This type of SCTM model has been applied to traffic flow on signalized intersection [11] by ignoring the signal control effect on traffic flow. In this research, we developed a dynamical model to determine the dynamic changing process of traffic flow on signalized intersection by switching linear dynamical systems (SLDS). SLDS are useful in describing dynamical phenomena, particularly structural changes over time. SLDS can be viewed as an extension of hidden Markov models (HMMs) [12]. Each HMM mode can be associated with the patterns of traffic flow transition such as on freeway [3] or signalized intersection condition [13]. The patterns or modes of traffic flow can be defined according to the characteristic of propagating shockwave on signal phases of which speed can be sampled from SFD. When the traffic density evolves according to the triangular FD, conditioned on an appointed mode sequence, the SLDS should ideally be consistent with the SCTM.

After constructing the SLDS model, the key point is to find the exact expressions for the approximate parameter distributions of the SFD. In our SLDS with time series data, the traffic state varies in time and is structured by the SFD conditioned on mode sequence. Assuming a certain framework with mode sequence and observations, the distributions of shockwave speed can be expressed by the posterior probability distribution, approximately found by the variational Bayesian method [14].

The organization of this paper is as follows. In Section 2, we will discuss our proposed approach for modeling traffic flow in signalized intersection. And in Section 3, the SFD is constructed by a local linear model with stochastic variables. Section 4 details the variational Bayesian learning method for the SFD's parameter distribution inference. Section 5 provides numerical verification for the model and algorithm presented in Sections 3 and 4. The model and algorithm

are validated using the next-generation simulation (NGSIM) data. Section 6 concludes the paper and briefly discusses future research directions. Appendix A is the derivation process of the posterior distribution of SFD which is proved to be the approximation distribution in Appendix B.

2. The Proposed Model for Traffic Flow on Signalized Intersection

2.1. The Minimal Representation of Road Segment. Consider the simple setting of a homogeneous roadway with signal control, as shown in Figure 1(a). The road segment between adjacent intersections can be divided into two simplest cells (link cells), where Cell(2) represents the road area, Cell(1) represents the intersection area, and Cell(1, 2) represents the signalized link shown in Figure 1(b). If the adjacent intersections are far away, the number of cells can be increased. But, at the same time, too much size of cell will reduce the efficiency of calculation. The cells' lengths must satisfy the constraint condition: $l_x/\Delta t \leq v_{\max}$, where v_{\max} denotes the maximum transit speed of the flow in the cell and Δt is time interval. The input and output traffic flow can be considered as the flow from or to a "boundary" cell that is named a "virtual" cell in our research, shown in Figure 1(b). The virtual cell can be defined with any length and capacity corresponding to the sending and receiving flow associated with signal control. For a better explanation, we define the traffic states located at the three sections as input, output, and transition traffic flow.

2.2. State Mode of Traffic Flow on Signalized Intersection. For the simplest signalized link case. The state mode of traffic flow can be defined by the shockwave propagation assumed to be located at the boundary between Cell(1) and Cell(2) on freeway [3]. Because of cyclic signal phase changes, the traffic flow wave propagating at a signal-controlled intersection has a cyclic pattern [13], as shown in Figure 2. In our research, we define the state mode of traffic flow by considering both the transition wave at the boundary between Cell(1) and Cell(2) and the boundary condition with cyclic changes. As indicated in Figure 2(a), at the beginning of a red light (assume there is no residual queue at the beginning of a cycle), the shockwave $\omega_b(1)$ is generated and propagates backward. In this period, the traffic

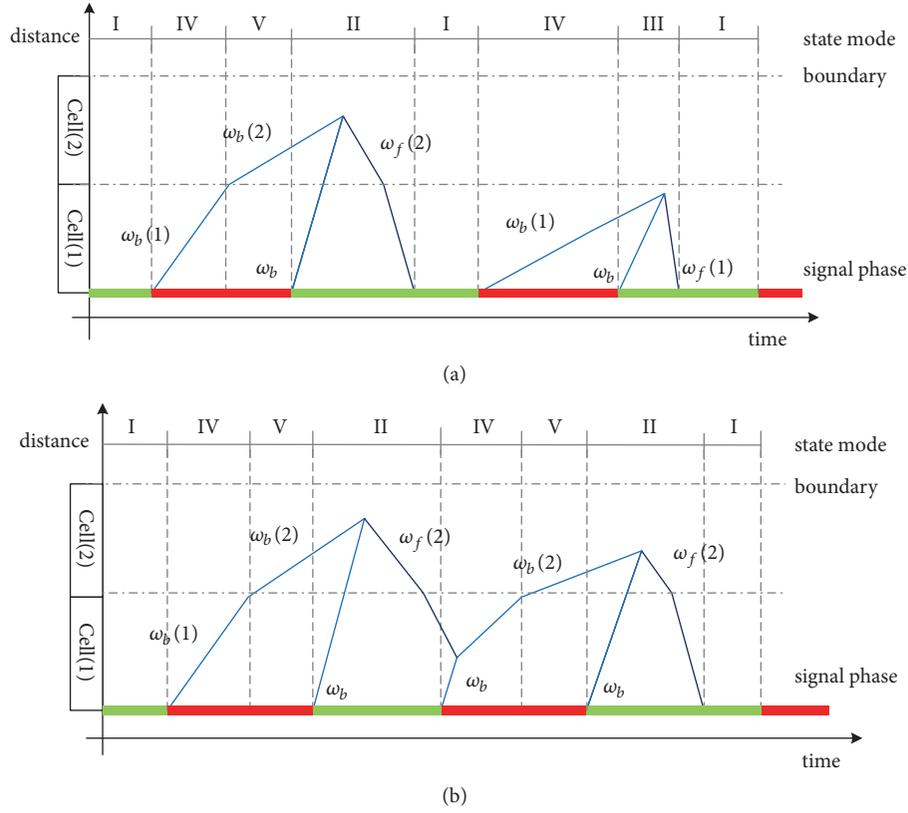


FIGURE 2: State mode of traffic flow corresponding to shockwave propagation: (a) without residual queues, (b) with residual queues.

state changes from the “free flow–free flow” in Cell(1) and Cell(2) to “Free flow-Congestion1” due to the queuing process in Cell(1) caused by the red light. After the shockwave $\omega_b(1)$ reaches the boundary between Cell(1) and Cell(2), the shockwave $\omega_f(2)$ which keeps propagating forward with input flow rate q_{in} is replaced by shockwave $\omega_b(2)$. And the traffic state changes to “Free flow-Congestion2” that is different from “Free flow-Congestion1” because of the direction of wave propagation. Once the green light is given, the queue begins to discharge with flow rate q_{max} . At the same moment, the discharge shockwave ω_b is generated in the stop-bar and propagates backward. The queue reaches its maximum length when the shockwave $\omega_b(2)$ meets the shockwave ω_b , and shockwave $\omega_f(2)$ is generated and propagates forward. Once the shockwave $\omega_f(2)$ reaches boundary between Cell(1) and Cell(2), where the shockwave $\omega_f(1)$ is generated and propagates forward, the queue discharge completes and the traffic state remains in “Free flow-Congestion2” until the shockwave $\omega_f(1)$ reaches the stop-bar. The traffic state returns to “Free flow- Free flow.” In this period, the transition wave is dominated by shockwave ω_b for most of the time which is the shockwave $\omega_f(2)$ in very short time due to the maximum output flow rate q_{max} .

In the red light period, if the shockwave $\omega_b(1)$ does not reach the boundary between Cell(1) and Cell(2), the traffic state will remain in “Free flow-Congestion1” until the end of the queue discharge, as shown in Figure 2(a). This situation

arises because the capacity of Cell(1) is very large or the input flow is particularly small resulting in a very short queue length.

If a queue does not fully discharge by the end of the cycle, a residual queue is formed, as shown in Figure 2(b). After the next cycle red phase starting, the shockwave $\omega_b'(1)$ propagates backward which has the same speed as the shockwave ω_b . Then, the other processes are just the same as those above.

The fact that the shockwave pattern is formation from cycle to cycle due to the signal control makes the shockwave identify traffic state modes on a signalized link. So, the state modes can be summarized by the shockwave pattern and boundary conditions with signal phase, as in Table 1.

2.3. State Mode-Based Model of Traffic Flow on Signalized Intersection. Consider a road segment with flow through its boundaries at time t . By combining FDs which express the mean flow rate as functions of mean traffic density, the numerical solution of the conservation laws is obtained by recursively computing the following:

$$\rho(x, t + \Delta t) = \rho(x, t) + \frac{\Delta t}{l_x} (q(x - 1, t) - q(x, t)), \quad (1)$$

where $\rho(x, t)$ denotes the mean traffic density in Cell(x) at time t , l_x is the length of the road segment, and Δt is time interval, while $q(x - 1, t)$ and $q(x, t)$ denote the mean flow

TABLE 1: State modes of traffic flow in two cells link for intersection arterial.

State mode	The shockwave pattern	The boundary condition	Signal phase	Traffic state
I	forward wave $\omega_f(2)$	forward wave $\omega_f(1)$ input flow rate q_{in}	green	Free flow-Free flow
II	backward wave ω_b	output flow rate q_{max} input flow rate q_{in}	green	Free flow-Congestion2
III	forward wave $\omega_f(2)$	output flow rate q_{max} input flow rate q_{in}	green	Free flow-Congestion1
IV	forward wave $\omega_f(2)$	output flow rate 0 input flow rate q_{in}	red	Free flow-Congestion1
V	backward wave $\omega_b(2)$	output flow rate 0 input flow rate q_{in}	red	Free flow-Congestion2

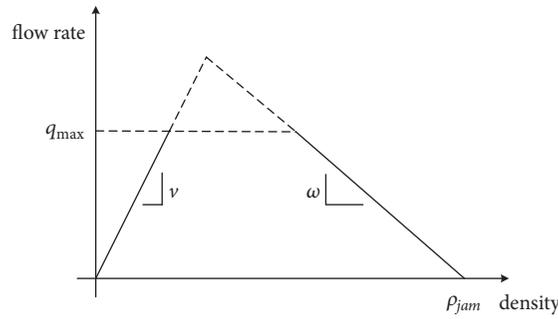


FIGURE 3: Triangular FD for the generalized CTM.

rate through the upstream and downstream boundaries of the link Cell($x-1, x$), respectively. The FD can then be written as

$$q(x, t) = F(\rho(x, t)). \quad (2)$$

In the simple case of a concave flux function, (2) could be given by CTM [7, 8]:

$$q(x, t) = \min \{v\rho(x-1, t), q_{max}, w(\rho_{jam} - \rho(x, t))\}, \quad (3)$$

where v is the forward speed, w is the backward wave speed, ρ_{jam} is the jam density, and q_{max} is the maximum flow. In this case, the FD is assumed to be a triangular function, as shown in Figure 3.

With the triangular FD, the traffic flow rate through the upstream or downstream boundaries is determined by the wave propagation's speed and direction (forward or backward) dependence on adjacent cell traffic densities. In this paper, we introduce a stochastic process to describe the stochastic wave propagation direction and define the SFD to compute distributions of the traffic state. So (1) is rewritten as

$$\begin{aligned} \rho(x, t + \Delta t) &= \rho(x, t) \\ &+ \frac{\Delta t}{l_x} (\mathcal{F}(\rho(x-1, t)) - \mathcal{F}(\rho(x, t))) \\ &+ e_t, \end{aligned} \quad (4)$$

where $\rho(x, t)$ is a stochastic variable denoting the traffic density in Cell(x) at time t , $\mathcal{F}(\rho(x, t))$ is the SFD, and e_t is process noise that is mode-specific with Gaussian distribution (with mean zero and covariance Σ).

Owing to cyclic signal phases and random arrival and departure of traffic flow, the traffic state dynamically changes from one mode set to others, as described in Section 2.2. Using (4), the dynamic evolution process of density can be modeled using a set of two linear systems, with which the parameter respectively is the traffic mode and density. This is known as a SLDS, as shown in Figure 4.

At each time t , in the first-level process, a switch traffic state mode $z_t = \{1, \dots, k, \dots, K\}$ describes which one in a set of linear dynamical systems (LDS) is to be used. In the second-level process, the transition dynamics of the continuous traffic density $\rho(t)$ are linear. Then, the switching-mode model is used to transform a CTM's nonlinear state equations into a set of piecewise linear equations, similar to SCTM for freeway [3, 9], and (4) becomes

$$\rho(t + \Delta t) = \rho(t) + \frac{\Delta t}{l_x} f_{z_t}(\rho(t)) + e_t(z_t) \quad (5)$$

$$\forall \rho(t) \in [0, \rho_{jam}]$$

$$f_{z_t}(\rho(t)) = A_{z_t} \rho(t) + B_{J, z_t} \rho_J + B_{Q, z_t} \mathbf{q}_{max} \quad (6)$$

$$e_t(k) \sim \mathcal{N}(0, \Sigma^k), \quad (7)$$

TABLE 2: Transition matrix of (6) for the state modes.

z_t	A_{z_t}	B_{J,z_t}	B_{Q,z_t}
I	$\begin{bmatrix} \omega_f & & & \\ \omega_f & -\omega_f(2) & & \\ & \omega_f(2) & -\omega_f(1) & \\ & & \omega_f(1) & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} 0 \\ & & & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} 0 \\ & & & \end{bmatrix}_{4 \times 4}$
II	$\begin{bmatrix} \omega_f & & & \\ \omega_f & & \omega_b & \\ & & -\omega_b & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} & & -\omega_b & \\ & & \omega_b & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} & & & -1 \\ & & & 1 \end{bmatrix}_{4 \times 4}$
III	$\begin{bmatrix} \omega_f & & & \\ \omega_f & -\omega_f(2) & & \\ & \omega_f(2) & & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} 0 \\ & & & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} & & & -1 \\ & & & 1 \end{bmatrix}_{4 \times 4}$
IV	$\begin{bmatrix} \omega_f & & & \\ \omega_f & -\omega_f(2) & & \\ & \omega_f(2) & & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} 0 \\ & & & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} 0 \\ & & & \end{bmatrix}_{4 \times 4}$
V	$\begin{bmatrix} \omega_f & & & \\ \omega_f & \omega_b(2) & & \\ & -\omega_b(2) & & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} & & -\omega_b(2) & \\ & & \omega_b(2) & \end{bmatrix}_{4 \times 4}$	$\begin{bmatrix} 0 \\ & & & \end{bmatrix}_{4 \times 4}$

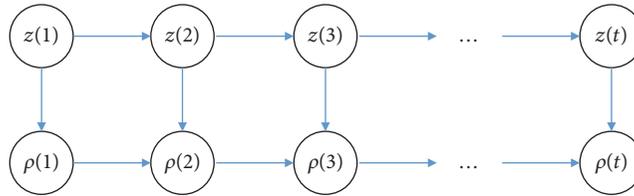


FIGURE 4: Graphical model of SLDS.

where $\rho(t)$ is the vector of traffic density states at time t with the dimension x , ρ_{jam} is the vector of jam densities, and \mathbf{q}_{max} is the vector of maximum flow rates.

When the conservation law is added to the noise in (5), the density must be subject to $[0, \rho_{jam}]$. The switch variable $z_t, z_t = \{1, \dots, k, \dots, K\}$, selects a single transition matrix from each of the available sets, $\{A_1, A_2, \dots, A_k\}$, $\{B_{J,1}, B_{J,1}, \dots, B_{J,k}\}$, and $\{B_{Q,1}, B_{Q,1}, \dots, B_{Q,k}\}$. The dynamics of z_t itself are Markovian, with transition probability:

$$p(z_t | z_{t-1}) \sim \pi_t. \quad (8)$$

Considering the SLDS model consisting of four cells indicated in Section 2.1, we specify the transition matrixes in (6) under each mode of SLDS as listed in Table 2. Note that the input flow rate is q_{in} and the upstream virtual cell density

at time t is $\rho_1(t) = q_{in}(t)\Delta t/l_x$ in $\rho(t)$, where $\omega_f = l_x/\Delta t$. Because the downstream virtual cell flow rate is zero in red light, in this model, the shockwave $\omega_b(1)$ cannot be derived directly.

3. Stochastic Fundamental Diagram (SFD)

In this section, we describe the SFD. According to (5)–(8), this stochastic process can be described by HMM with the stochastic variables of traffic density and traffic state mode. From time t to $t + \Delta t$, the traffic density of Cell(x) is $\rho^{z_t}(x, t)$ to $\rho^{z_{t+\Delta t}}(x, t + \Delta t)$, and the corresponding flow rate is $q(x-1, t)$ and $q(x, t)$ through the Cell(x). If FD appears nonlinear and flux such as in Figure 5, we must define the SFD function $\mathcal{F}(\rho^{z_t}(x, t))$ in (4) to replace (2). If SFD is not analytically

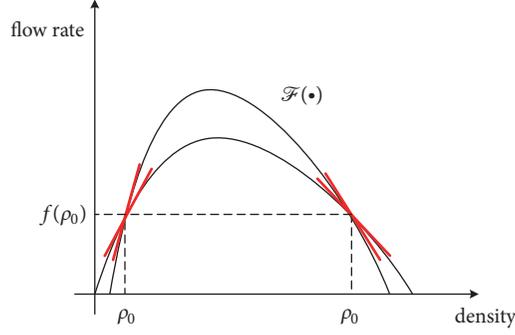


FIGURE 5: Local model to approximate the SFD.

tractable and cannot be modeled by a function completely, an alternate approach is to apply the linear regression model locally to approximate the real SFD [15].

Given a nonlinear function $\mathcal{F}(\bullet)$ that could be any curve, as shown in Figure 5, we construct a first-order Taylor series approximation in any local area of $\mathcal{F}(\bullet)$ by given mode $z_t = k$ as

$$\mathcal{F}(\rho^k(x, t)) \approx q(\rho_0^k(x, t)) + \omega^k(\rho^k(x, t) - \rho_0^k(x, t)), \quad (9)$$

where $q(\rho_0^k(x, t))$ is the SFD's function value of $\rho_0^k(x, t)$, ω^k is the wave's propagation speed, and $\omega^k = \nabla \mathcal{F}(\rho^k(x, t))|_{\rho^k(x, t) = \rho_0^k(x, t)}$.

The SFD can be approximated by n tangents passing through point $(\rho_0^k, q(\rho_0^k))$ constrained by the limit of the maximum flow rate q_{\max} , shown in Figure 5, so that

$$\mathcal{F}(\rho^k(x, t)) \approx \begin{cases} f_n(\rho^k(x, t)) = [q(\rho_0^k) + \omega^k(\rho^k(x, t) - \rho_0^k)]_n \\ q_{\max} \end{cases}, \quad (10)$$

where $f_n(\rho^k(x, t))$ denotes the local approximation function and $[\bullet]_n$ implies n numbers. If ω^k is a stochastic variable subjected to a certain distribution, a more general definition of the SFD is given as follows.

Definition 1. Given the stochastic function variable $\rho^k(x, t)$ and stochastic parameter variables ω^k belonging to the K th mode, the variables $[\omega^k]_n$ are subject to independent and identical distribution (IID) \mathcal{R} . The SFD can be written as

$$\mathcal{F}(\rho^k(x, t)) \approx \begin{cases} q(\rho_0^k) + [\omega^k]_n(\rho^k(x, t) - \rho_0^k) \\ q_{\max} \end{cases}, \quad (11)$$

where

$$[\omega^k]_n \sim \mathcal{R}. \quad (12)$$

With Definition 1 applied, SFD is the distributions of parameter ω^k . When the wave propagates forward, the density data

group is $\rho^k \in [0, \rho_c)$, where $0 < \rho_c < \rho_J$. Let $\omega_f^k > 0$; then if the Taylor series is expanded at point $(0, 0)$, meaning $\rho_0^k(x, t) = 0$ and $f_n(\rho_0^k(x, t)) = 0$, the SFD is approximated by n numbers of rays forming at the point $(0, 0)$, as shown in Figure 6(a). When the wave propagates backward, the density data group is $\rho^k \in (\rho_c, \rho_J]$, where $0 < \rho_c < \rho_{jam}$. Let $\omega_b^k > 0$, so the Taylor series expansion at point $(\rho_{jam}, 0)$ is $\rho_0^k(x, t) = \rho_{jam}$ and $f_n(\rho_0^k(x, t)) = 0$. The SFD is approximated by n rays forming at point $(\rho_{jam}, 0)$, as shown in Figure 6(a). Then, SFD can be written as

$$\mathcal{F}(\rho^k(x, t)) \approx \begin{cases} [\omega_f^k]_n \rho^k(x, t) & \rho^k \in [0, \rho_c) \\ [\omega_b^k]_n (\rho_{jam} - \rho^k(x, t)) & \rho^k \in (\rho_c, \rho_J] \\ q_{\max} \end{cases} \quad (13)$$

So (3) can be recognized as a special solution of (13). And (13) can be regarded as the family of triangular FDs with different values of density by the added parameter $[\omega_f^k, \omega_b^k]_n$. If we assume that the parameters $[\omega_f^k]_n, [\omega_b^k]_n$ are subject to independent and identical Gaussian distribution $\mathcal{N}(\omega_f^k; \mu_f^k, \varsigma_f^k), \mathcal{N}(\omega_b^k; \mu_b^k, \varsigma_b^k)$, as shown in Figure 6(b), then (6) can be sampled from $\mathcal{F}(\rho(t))$ in (13) condition on $z_t = k$:

$$f_{z_t}(\rho(t)) \sim \mathcal{F}(\rho^k(t) | z_t = k). \quad (14)$$

4. Variational Bayesian (VB) Learning for SFD

Once the traffic flow on freeway has been modeled by SCTM with FD [9, 16], the parameters ω_f^k, ω_b^k can be obtained from the maximum likelihood estimation (MLE) method by giving the independent observation sequence $\{\rho(t)\}_{t=1:T}$ and state mode sequence $\{z_t\}_{t=1:T}$ [17], which equals the parameter calibration on freeway condition [9, 16]. But we want to obtain the distribution of parameters ω_f^k, ω_b^k in (12) on the basis of Definition 1 for the traffic flow on signalized intersection, which can be computationally approximated as the posterior distribution by the variational Bayes (VB) learning method for SLDS [18].

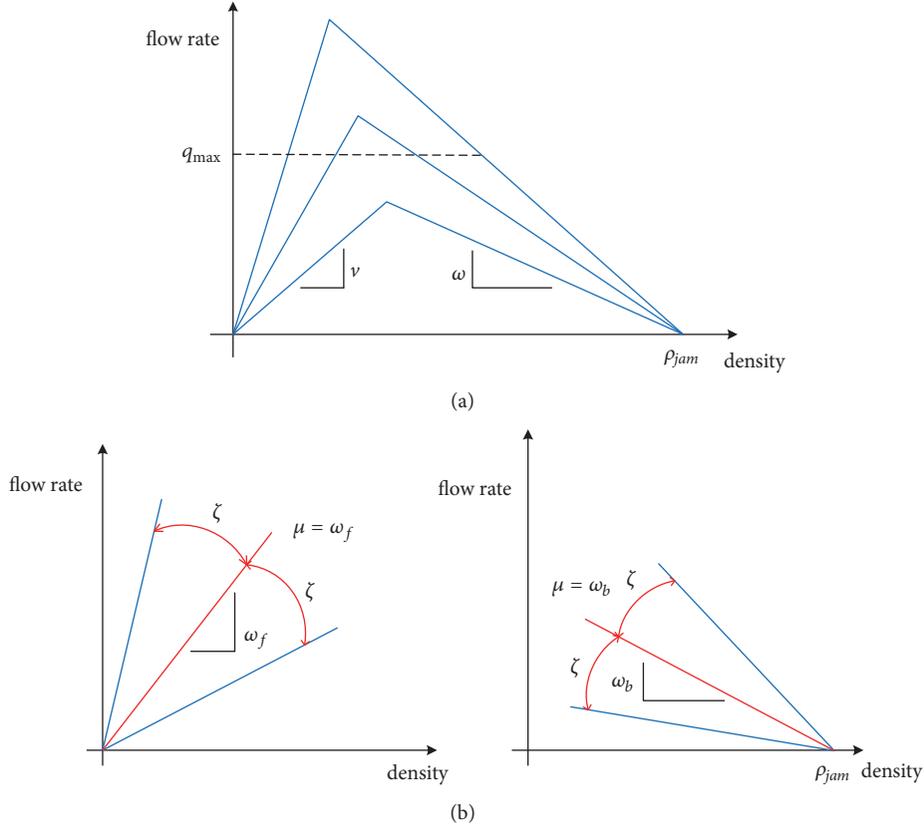


FIGURE 6: SFD defined by family of the triangular FDs: (a) family of the triangular FDs, (b) the parameters of SFD defined by Gaussian distribution: any triangular FD in the family can be calculated by the Gaussian distribution.

The basic idea of variational Bayesian learning is to simultaneously approximate the intractable joint distribution over both hidden states and parameters of discrete graphical models with a simpler distribution, usually by assuming the hidden states and parameters are independent [18]. In our problem, the hidden states of the discrete graphical model are correspondingly the traffic density and traffic state mode of SLDS. Given the independent traffic density sequence $\{\rho(t)\}_{t=1:T}$ and state mode sequence $\{z_t\}_{t=1:T}$, the unknown variables of SLDS are parameters ω_f^k, ω_b^k and covariance Σ^k . Let the training dataset be as follows:

$$D = \{(\bar{\rho}^k, \rho^k) \mid z_t = k\}, \quad (15)$$

where $(\bar{\rho}^k, \rho^k)$ are the $(T-1)$ dimension vectors: $\rho^k = \{\rho(t)\}_{t=1:T-1}$ and $\bar{\rho}^k = \{\rho(t)\}_{t=2:T}$.

For the parameter set $\Theta = \{\omega_f^k, \omega_b^k\}$, the posterior of parameter $p(\Theta \mid D)$ is computed by Bayesian expression as

$$p(\Theta \mid D) \propto p(D \mid \Theta) p(\Theta), \quad (16)$$

which can be rewritten as

$$p(\omega^k \mid D, \Sigma^k) \propto p(D \mid \omega^k, \Sigma^k) p(\omega^k), \quad (17)$$

where $p(\Theta)$ or $p(\omega^k)$ is the prior of parameters ω_f^k, ω_b^k and $p(D \mid \Theta)$ or $p(D^k \mid \omega^k, \Sigma^k)$ is the likelihood and identical

to the one in MLE [17]. If the prior of ω^k is selected as the Gaussian, and $\{\rho(t)\}_{t=1:T}$ represents the IID of the Gaussian, the conjugate prior of the covariance Σ^k is the inverse-Wishart distribution, denoted by IW [19].

Considering (6), the matrix A_{z_t}, B_{J, z_t} is composed of the same parameter ω^k by different modes z_t . The matrix A_{z_t}, B_{J, z_t} can be converted to a vector $vec \bar{A}_k$. In Table 2, we find that the matrix A_{z_t}, B_{J, z_t} is the sparse matrix and $vec \bar{A}_k$ is also a sparse matrix. So, we consider the automatic relevance determination (ARD) which encourages driving the matrix elements to zero if their presence is not supported by the model or data [14]. Then, the parameter priors are given by

$$p(vec \bar{A}_k) = \prod_i \mathcal{N}(\omega_j^k; 0, \alpha^{-k} I_n) \quad (18)$$

$$p(\Sigma^k) = IW(\Sigma^k; n_0, S_0). \quad (19)$$

Finally, the parameter posterior is (detailed in Appendix A)

$$p(\omega^k \mid D, \Sigma^k) = \mathcal{N}(\omega_j^k; \mathbf{N}, \Lambda). \quad (20)$$

Remark 2. Given the parameter priors as in (18), the covariance prior as in (19), and the likelihood as in (A.6), the parameter $[\omega_f^k]_n, [\omega_b^k]_n$'s distribution \mathcal{R} in Definition 1 has the approximation form in (20) (proof in Appendix B).

1 Initialization:

Given $\{\rho(t)\}_{t=1:T}$, mode sequence $\{z_t\}_{t=1:T}$, and a previous set of dynamic parameters $\Theta = \{\omega^k\}$ and Σ^k

2 Data updates:

Compute the expected sufficient statistics $D_k = \{\mathbf{Q}_k, \bar{\mathbf{P}}_k\}$ as in Eq. (A.3)

3 Iteration:

(a) Construct Σ_0^k given α^k as in Eq. (A.5)

(b) Sample the dynamic matrix $\text{vec}\bar{A}_k \mid \Sigma_0^k, \alpha^k$ with the distribution as in Eq. (A.9)

(c) Sample ARD precision parameters $\alpha^k \mid A_k$ with the distribution as in Eq. (A.16)

(d) Compute the sufficient statistic $S_{D_k}^k$ with Eq. (A.15)

(e) Sample process noise covariance $\Sigma^k \mid A_k$ with the distribution as in Eq. (A.14)

4 End

ALGORITHM 1: Parameter Gibbs Sampling.

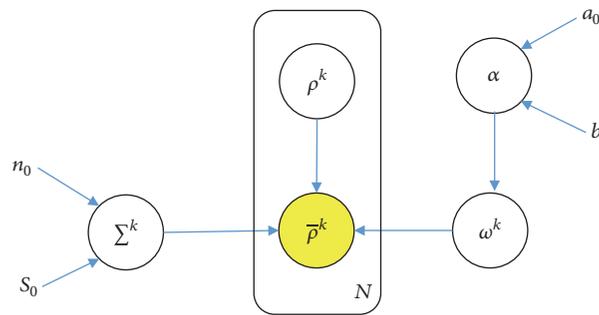


FIGURE 7: The interdependencies of the parameters, variables, and hyperparameters of graphical model.

As the optimal form of the parameter distribution is given, the concrete distribution can be obtained by Gibbs Sampling with conditions on observation data [20, 21]. The interdependencies among model parameters, variables, and hyperparameters are shown in Figure 7. Here, we use the Parameter Gibbs Sampling in Algorithm 1 to obtain the approximated posterior distributions using (20).

5. Empirical Verification

5.1. Data Introduction. The data utilized in this research was collected on a segment of Peachtree Street, in Atlanta, Georgia, USA, on November 8th, 2006, as part of the Federal Highway Administration (FHWA) NGSIM [22]. The data includes detailed individual vehicle trajectories with time and location, vehicle velocity, and headway (distance) between vehicles, from which the flow-density relationship can be calculated. The Peachtree Street segment is approximately 2,100 feet in length, with five intersections and two or three through-lanes in each direction. Intersections 1, 2, 3, and 5 are signal-controlled intersections, while Intersection 4 is a stop sign-controlled intersection. The Peachtree Street data consist of two 15-minute time periods, 12:45 p.m. to 1:00 p.m. and 4:00 p.m. to 4:15 p.m. The traffic volume in 4:00 p.m. to 4:15 p.m. is larger than the volume in 12:45 p.m. to 1:00 p.m., so our study focused on the segment between Intersections 2 and 3 in 4:00 p.m. to 4:15 p.m. Intersection 2 is a two-phase signal-controlled intersection (green, red, and amber lights).

In our study's time period, the cycle length is constant (102s) with the g/C ratio being 0.43.

5.2. Observations and Empirical Distributions. As shown in Figure 8, the segment between Intersections 2 and 3 is divided into two cells, with lengths of 167 feet and 170 feet for Cell(1) and Cell(2), respectively. The average traffic flow rate and density in each cell are calculated in every 5s interval. The average density and the jam density in each cell can be better estimated from the definition "vehicles per distance." Average volume of the traffic flow can be calculated for vehicles through the boundary. The traffic state sequence, cell density, and flow rate are shown in Figure 9. "Input" is the traffic flow rate in the upstream segment boundary and "Output" is the traffic flow rate in the downstream segment boundary. "Trans" is the traffic flow rate from Cell (2) to Cell (1).

The widely scattering flow-density relationship for the selected road segment, as shown in Figure 10, is obviously not tractable to be modeled by a function completely. We use the flow-density data to fit the Gaussian shockwave speed empirical distribution (the red line) shown in Table 3 and Figure 11. To compare with the triangular FD, we use the flow-density data to fit the linear model in (3) with $\rho_{jam} = 0.14$ (veh/feet), with the results shown in Table 4 and Figure 12.

To test the assumption of Gaussian shockwave speed empirical distribution statistically, the following hypothesis test by using t -test is conducted for each of the traffic state modes, as shown in Table 5. This test is used to determine if there is evidence that the shockwave speed comes from a

TABLE 3: Fitted Gaussian distribution of shockwave speed (feet/s).

	Mode I	Mode II	Mode IV	Mode V
	$\omega_f(1)$	ω_b	$\omega_f(2)$	$\omega_b(2)$
Mu	14.5422	4.11485	21.6132	3.64132
Mu-Std.err.	1.97247	0.586173	1.24654	0.452152
Sigma	8.36848	1.94412	5.9782	2.07202
Sigma-Std.err.	1.45677	0.446011	0.911666	0.331789

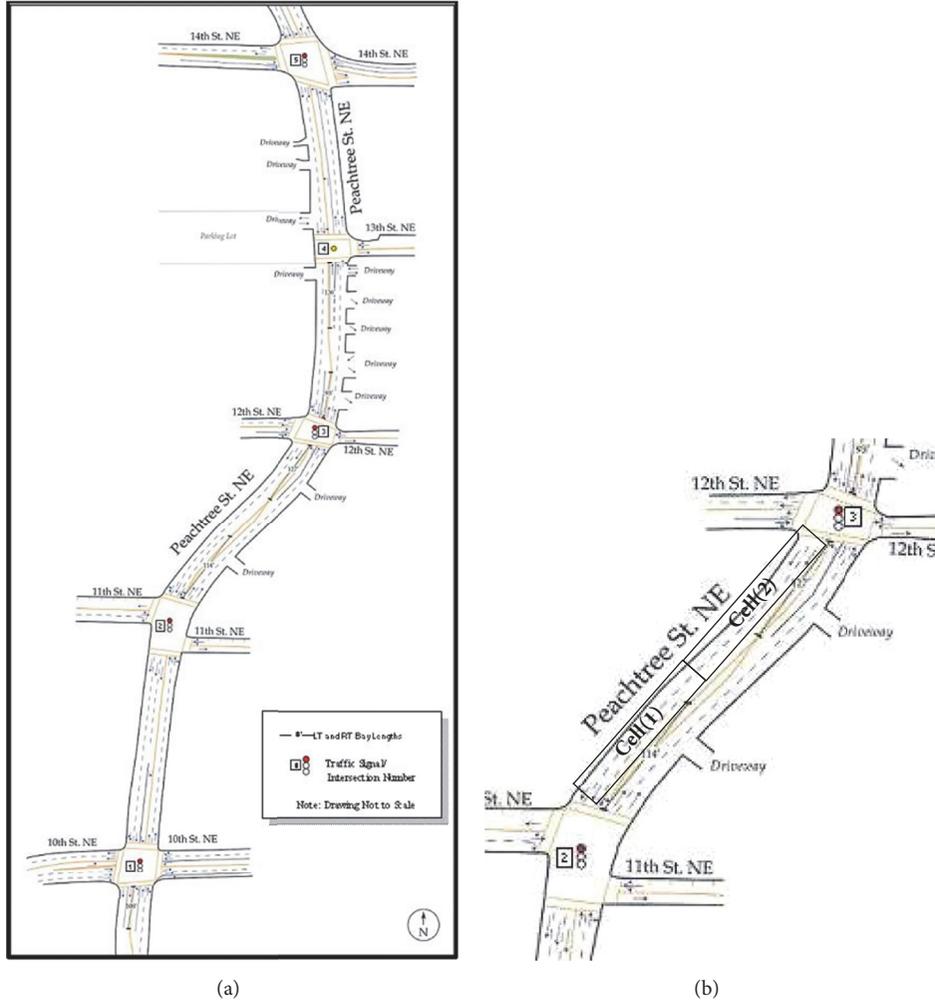


FIGURE 8: Selected target sections of NGSIM: (a) Peachtree Street in NGSIM, (b) the cell representation of the segment between Intersections 2 and 3.

Gaussian distribution with mean μ_0 and unknown variance. Using a confidence level of $\alpha = 0.05$, if the result is $P < 0.05$, the test rejects the null hypothesis H_0 and $P > 0.05$ otherwise. The ci is confidence interval for the true population mean.

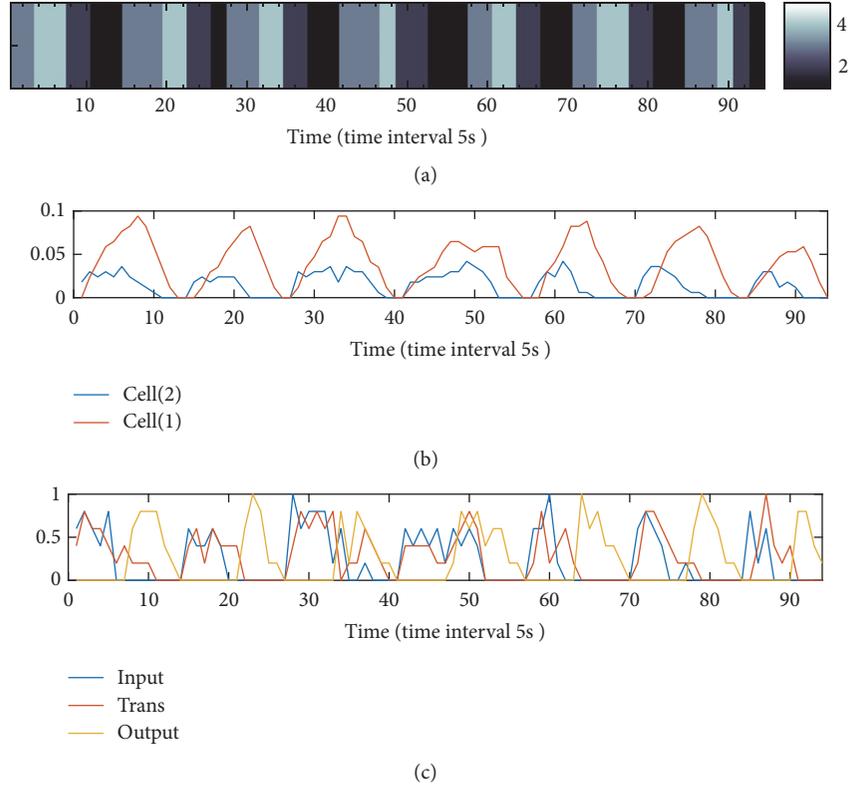
5.3. Numerical Verification. We examine the posterior distribution of the wave speed to analyze the SFD by giving the state sequence and cell density, as shown in Figure 9. Here, we select (13) expansion for SFD. We use $\mathcal{G}(a_0, b_0)$ as the precision parameter α_n^k of the wave priors in (18) and $IW(n_0, S_0)$ for the prior of covariance Σ^k in (19). We choose

the settings $a_0 = 1, b_0 = 0.01$ and $n_0 = 10, S_0 = 10I_4$. The results of the parameter posterior are listed in Table 6 and the posterior PDF (green line) are shown in Figure 11.

To the fitted empirical distributions, the posterior distributions can be the approximate distribution for the shockwave speed. And the means of inference posterior distributions are also included in the ci of hypothesis test. In Figure 13, the triangular FD shockwave speeds in every state mode group are almost equal to the mean of the SFD shockwave speed's distribution. Because we are given SFD shockwave speed's distribution, we can obtain the infinite lines whose

TABLE 4: Fitted triangular FD with wave speed (feet/s).

	Mode I	Mode II	Mode IV	Mode V
	$\omega_f(1)$	ω_b	$\omega_f(2)$	$\omega_b(2)$
shockwave speed	10.25	5.774	21.77	4.883
RMSE	0.1014	0.1754	0.1665	0.1944

FIGURE 9: The NG data for verification (time interval 5s): (a) state sequence z_t , (b) the cell density, and (c) the flow rate on three boundaries.

slope is the wave speed. The family line of SFD is a fan-shaped area that encompasses all the data according to the probability.

In Figure 14, the normal quantile-quantile of the NG wave speed versus the sample data, which are the 50 shockwave speeds sampled from the posterior, are displayed. The plot produces an approximately straight line, suggesting that the sample speed and the NG speed follow the same distribution.

It displays the likelihood $p(\mathbf{q}, \boldsymbol{\rho}^z \mid z, \Theta) = p(\mathbf{q}(t), \mathcal{F}(\boldsymbol{\rho}^k(t) \mid z_t = k), \Sigma^k)_{t=1:T}$ by using the inference result of parameter Θ and covariance Σ^k in Figure 15, and the flow-density is the same as that shown in Figure 10(b). We find that most of the flow-density data is distributed in the region in which the likelihood is the maximum one.

The traffic flow rate transit from Cell(2) to Cell(1) is decided by the adjacent cell's density and the state mode in SLDS with SFD or SCTM with calibrated FD [9]. By using the cell's density and state mode of NGSIM data in Figure 9, 100 times' random simulations are carried out to calculate the transition traffic flow rate where the shockwave speeds are sampling from SFD in order of traffic mode sequence in

each simulation. The Boxplot of transition traffic flow rate, removing the value of zero, in each time interval of all SLDS model simulations is shown in Figure 16 with the NGSIM data (red line) and simulation result of SCTM with calibrated FD (Blue line). The NGSIM data (red line) and simulation result of SCTM are included in the range of the SLDS model result which can be seen as a special simulation with SFD. The SLDS model with SFD is a more large-scale description of the dynamic characteristics of traffic flow than SCTM model with FD.

6. Conclusion, Limitations, and Future Research

In this work, first, the traffic flow on signalized intersection is modeled by SLDS model with SFD which is defined by the distributions of shockwave speed. Then, variational Bayesian learning method is adopted to compute the distributions of SFD parameter to approximate the experimental distributions of shockwave calculated by the observed flow-density data. It is shown that SFD can be more helpful to capture

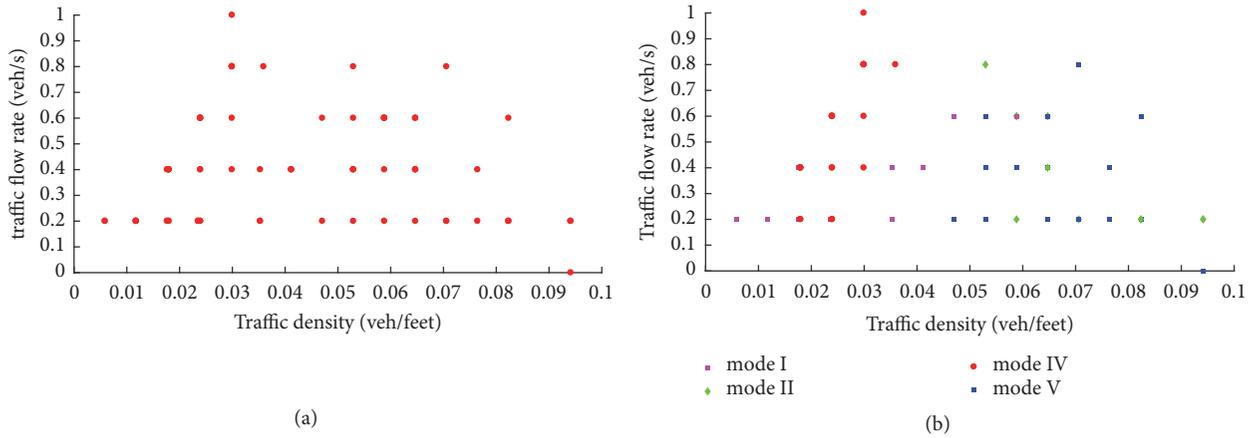


FIGURE 10: The flow-density relationship: (a) without condition on the mode, (b) with condition on the mode.

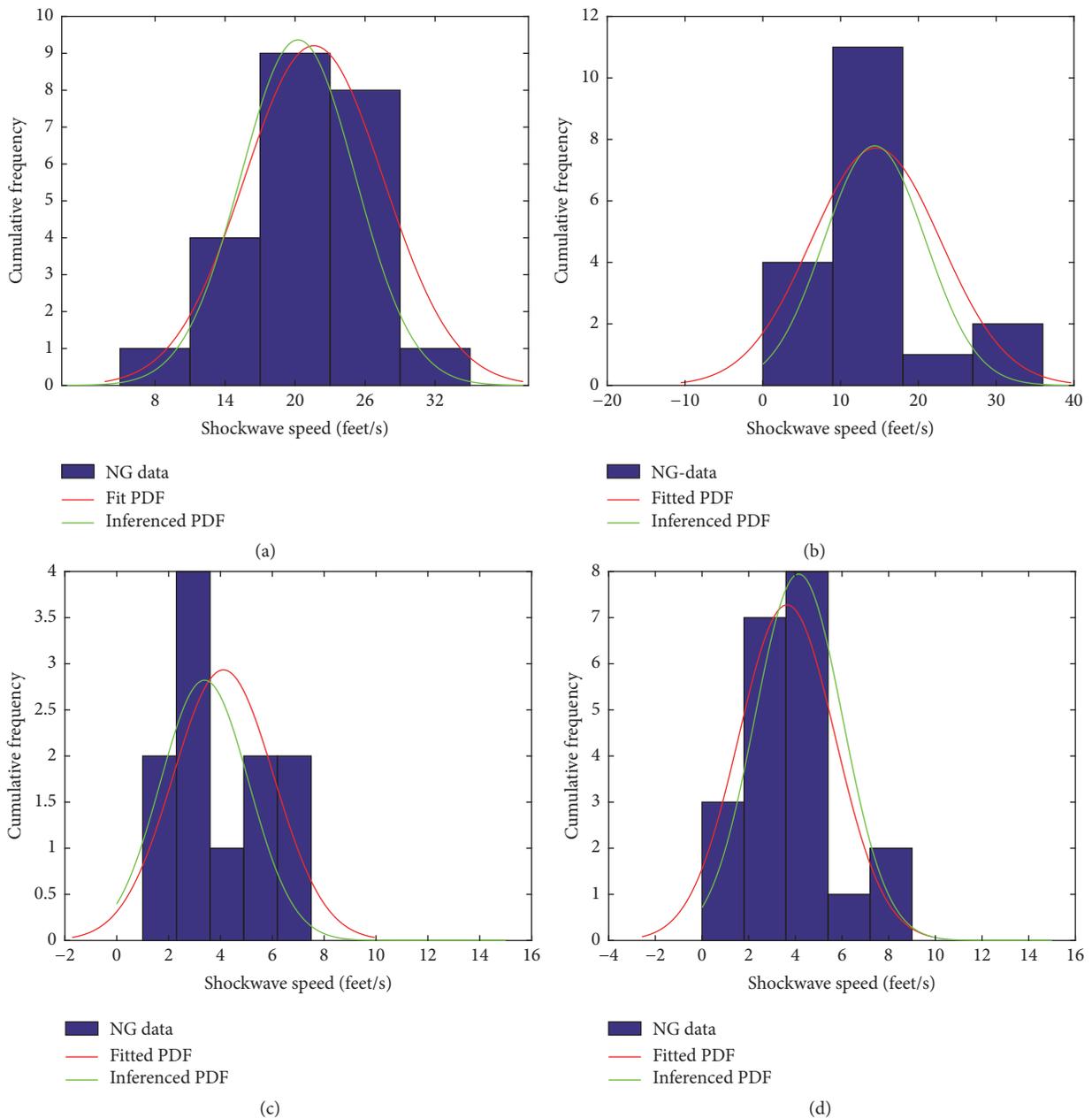


FIGURE 11: Fitted Gaussian and posterior distributions of wave speed (feet/s): (a) the wave $\omega_f(1)$ of mode I, (b) the wave ω_b of mode II, (c) the wave $\omega_f(2)$ of mode IV, and (d) the wave $\omega_b(2)$ of mode V.

TABLE 5: The hypothesis test for assumption of Gaussian shockwave speed distribution (feet/s).

	Mode I	Mode II	Mode IV	Mode V
	$\omega_f(1)$	ω_b	$\omega_f(2)$	$\omega_b(2)$
Average speed μ_0	14.5422	4.11485	21.6132	3.64132
Null hypothesis H_0	$\mu = \mu_0$	$\mu = \mu_0$	$\mu = \mu_0$	$\mu = \mu_0$
Alternative hypothesis H_1	$\mu \neq \mu_0$	$\mu \neq \mu_0$	$\mu \neq \mu_0$	$\mu \neq \mu_0$
Confidence level α	0.05	0.05	0.05	0.05
P	0.9991	0.9936	0.9980	0.9977
ci	[10.38, 18.70]	[2.81, 5.46]	[19.03, 24.19]	[2.70, 4.58]
Result	Not rejected	Not rejected	Not rejected	Not rejected

TABLE 6: Fitted Gaussian and posterior distribution of wave speed (feet/s).

	Mode I	Mode II	Mode IV	Mode V
	$\omega_f(1)$	ω_b	$\omega_f(2)$	$\omega_b(2)$
Mu-fit	14.5422	4.11485	21.6132	3.64132
Mu-posterior	14.3919	3.3856	20.9356	4.1440
ci	[10.38, 18.70]	[2.81, 5.46]	[19.03, 24.19]	[2.70, 4.58]
Sigma-fit	8.36848	1.94412	5.9782	2.07202
Sigma-posterior	6.53012	1.70941	4.90917	1.88689

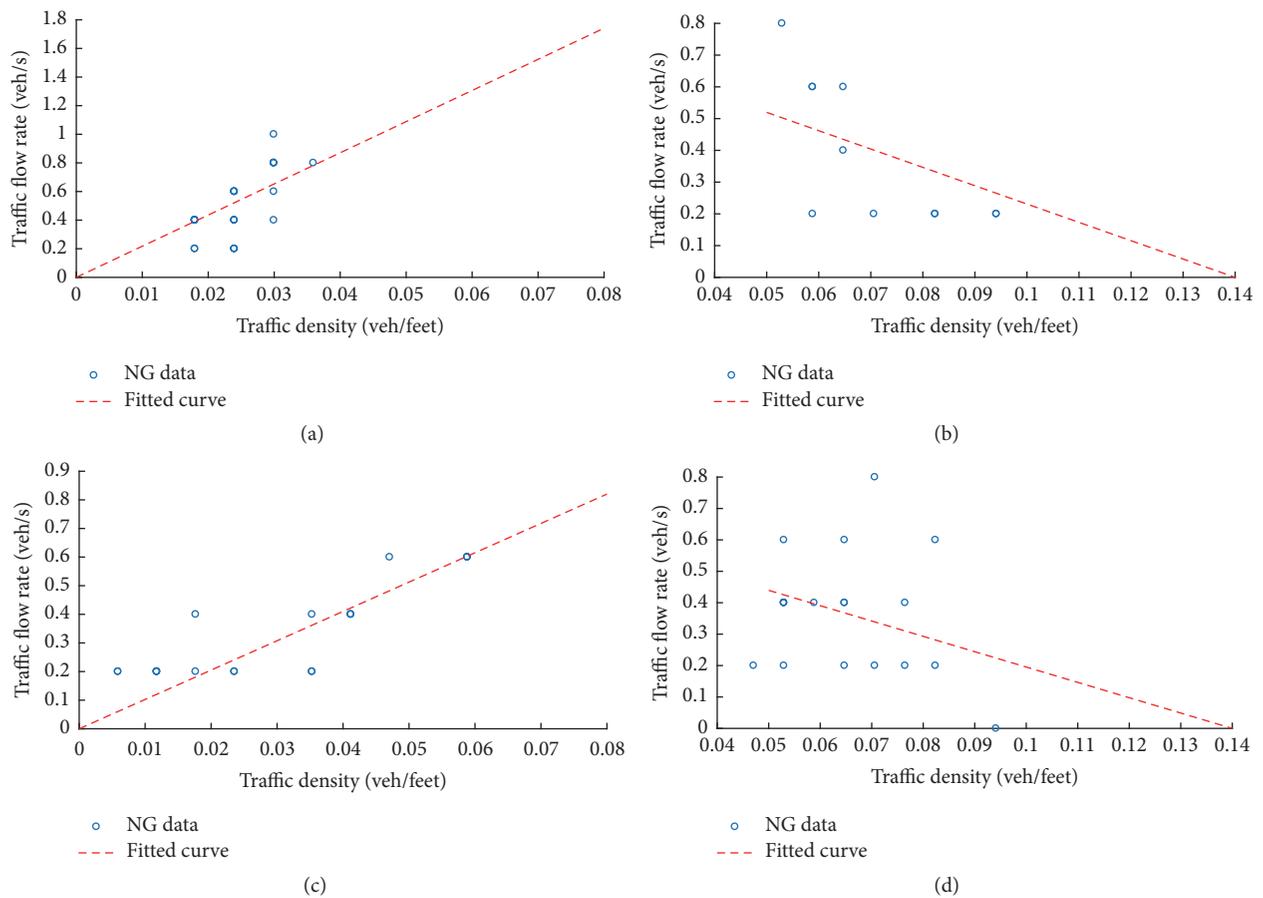


FIGURE 12: Fitted line of the four state modes' FD: (a) the wave $\omega_f(1)$ of mode I, (b) the wave ω_b of mode II, (c) the wave $\omega_f(2)$ of mode IV, and (d) the wave $\omega_b(2)$ of mode V.

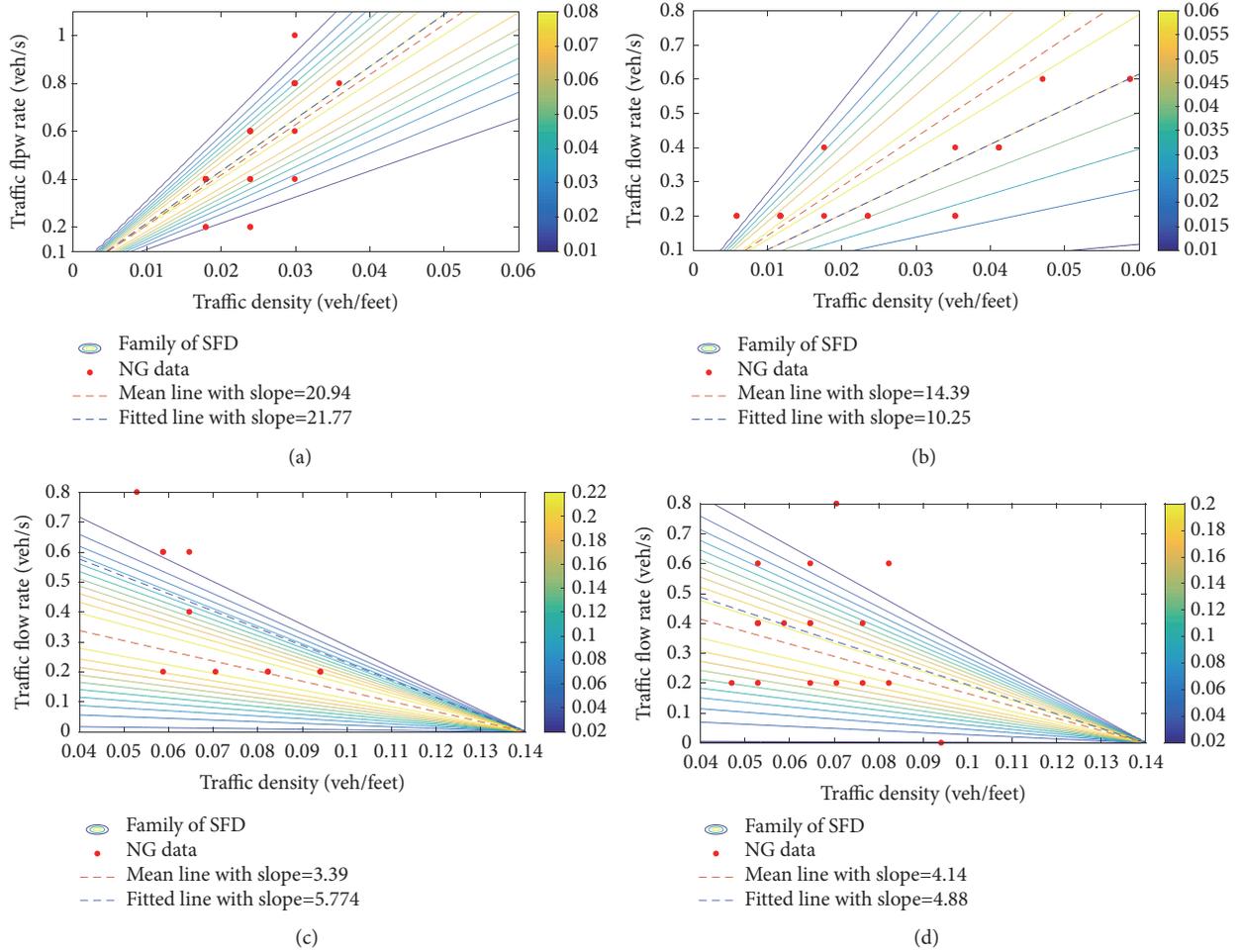


FIGURE 13: Fitted line of FD and family line of SFD: (a) the wave $\omega_f(1)$ of mode I, (b) the wave ω_b of mode II, (c) the wave $\omega_f(2)$ of mode IV, and (d) the wave $\omega_b(2)$ of mode V. The color bar represents the probability of a wave sampling from SFD. The color from deep to shallow indicates the probability from low to high.

the main features of the observed wide scattering of the flow-density data than FD tested by traffic flow data on Peachtree Street from the NGSIM program. With the wave speed sampled from SFD, the SLDS could describe the dynamic characteristics of traffic flow and be used to the maximum likelihood estimation for traffic density or flow rate.

Furthermore, SFD in this paper is also suitable for freeway traffic flow on the condition that the SLDS's traffic state mode is defined by freeway traffic flow such as in the SCTM [3, 9]. Once obtained, SFD may be used as an input together with the boundary conditions to determine the dynamical evolution of the traffic flow on signalized intersection. So SFD proposed in this paper could be applied to model-based signal control optimization [23].

One of the limitations of this model is that it does not consider scenarios in which the road segment in the intersection arterial is divided into straight lane group and turning lane group, which could potentially lead to the lane blockage phenomenon [24]. The model should consider the lane groups by expanding them to the form of a network [8]. And the traffic state mode should consider the traffic

flow spillover [13]. Another key point is that when using the Bayesian framework to infer SFD, we assume sufficient statistics equal to the observation, including the state mode sequence, density sequence, and cell boundary flow rate sequence. However, in practice, it is not easy to directly obtain the state mode sequence and the density sequence. Therefore, for SLDS model, we could infer the state mode and density which are set to be the latent variables when the traffic flow rate at the upstream and downstream boundaries of the road can be set to observation data, same as in the Hybrid Systems [24]. In this procedure, SFD could be automatically inferred by variational Bayesian learning method just like the automatic calibration [16].

Appendix

A. Posteriors of SFD Parameters

Considering (5)–(7), these equations are transformed as

$$\mathbf{Q}_{z_t} = \mathbf{A}_{z_t} \mathbf{P}_n + e_t(z_t), \quad (\text{A.1})$$

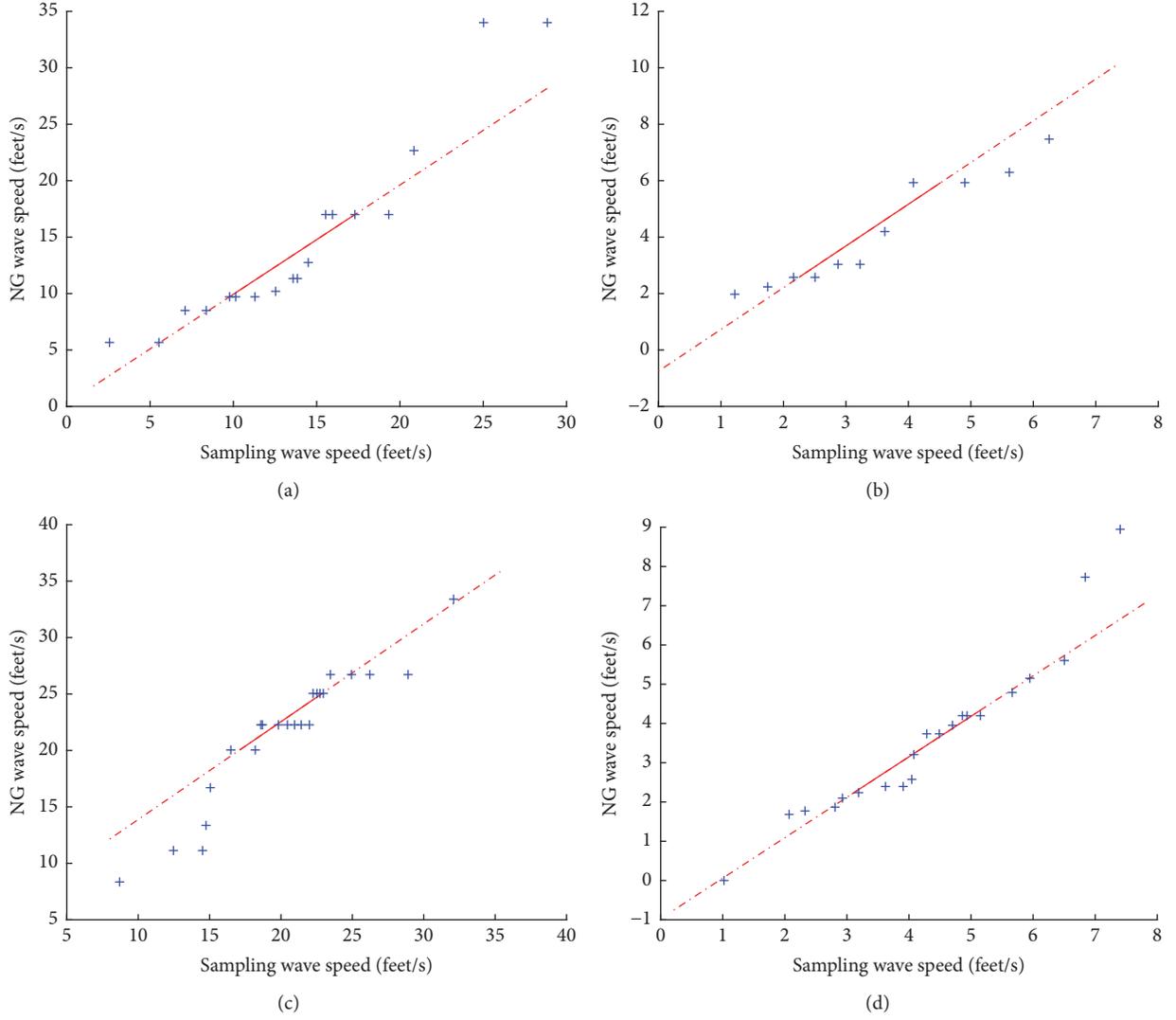


FIGURE 14: Normal quantile-quantile plot of wave speed: (a) the wave $\omega_f(1)$ of mode I, (b) the wave ω_b of mode II, (c) the wave $\omega_f(2)$ of mode IV, and (d) the wave $\omega_b(2)$ of mode V.

where

$$\mathbf{Q}_{z_t} = \frac{l_x}{\Delta t} [\boldsymbol{\rho}(x, t + \Delta t) - \boldsymbol{\rho}(x, t)] - B_{\mathbf{Q}_{z_t}} \mathbf{q}_{\max}$$

$$\mathbf{P}_n \in \left\{ [\boldsymbol{\rho}(x, t)]_n, [\boldsymbol{\rho}_J - \boldsymbol{\rho}(x, t)]_n, \right. \quad (\text{A.2})$$

$$\left. \left[\begin{array}{c} \boldsymbol{\rho}(x, t)_r \\ (\boldsymbol{\rho}_J - \boldsymbol{\rho}(x, t))_{n-r} \end{array} \right]_n \right\},$$

where \mathbf{P}_n is the n^{th} dimension vector and n is the number of cells. Let $\text{vec} \bar{A}_k = A_k(:)$; (A.1) can equivalently be represented as

$$\mathbf{Q}_{z_t} = \begin{pmatrix} \mathbf{P}_1 \cdots \mathbf{P}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{P}_n \cdots \mathbf{P}_n \end{pmatrix} \text{vec} \bar{A}_{z_t} + \mathbf{e}_t(z_t) \quad (\text{A.3})$$

$$= \bar{\mathbf{P}}_{z_t} \text{vec} \bar{A}_{z_t} + \mathbf{e}_t(z_t).$$

Given $D = \{(\bar{\boldsymbol{\rho}}^k, \boldsymbol{\rho}^k) \mid z_t = k\}$ and mode sequence $\{z_t\}_{t=1:T}$, we could obtain the set $D_k = \{\mathbf{Q}_k, \bar{\mathbf{P}}_k\}$.

Assume we place a Gaussian prior:

$$p(A_k) = \prod_i \mathcal{N}(\omega_j^k; 0, \Sigma_0^k) \quad (\text{A.4})$$

$$\Sigma_0^k = \text{diag}(\alpha_1^k, \alpha_2^k, \dots, \alpha_n^k)^{-1}. \quad (\text{A.5})$$

The likelihood model in (17) is

$$p(D \mid A_k, \Sigma^k) = \frac{1}{|2\pi\Sigma^k|^{N_k/2}} \cdot \exp\left(-\frac{1}{2} \left(\sum_i (\mathbf{Q}_k^i - \bar{\mathbf{P}}_k^i \text{vec} \bar{A}_k)^T \Sigma^{-k} (\mathbf{Q}_k^i - \bar{\mathbf{P}}_k^i \text{vec} \bar{A}_k) \right)\right). \quad (\text{A.6})$$

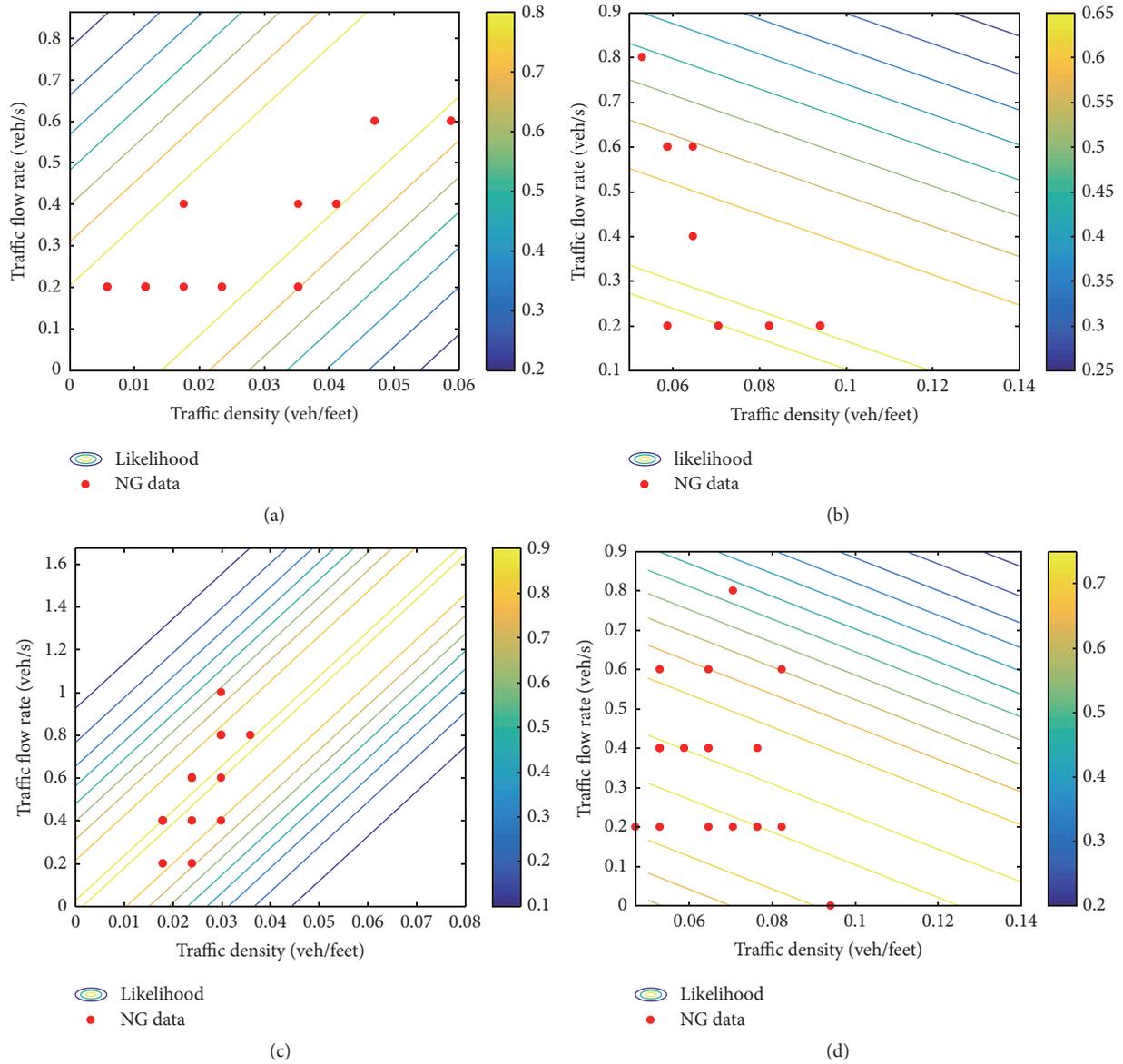


FIGURE 15: Likelihood of traffic flow density conditioned on traffic modes: (a) mode I, (b) mode II, (c) mode IV, and (d) mode V. The color bar represents the probability of likelihood. The color from deep to shallow indicates probability from low to high in [0, 1].

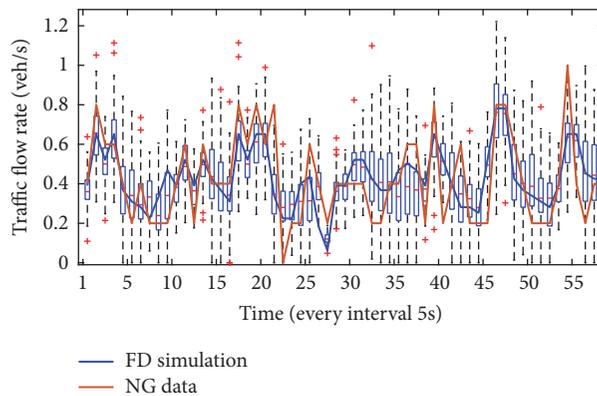


FIGURE 16: NGSIM transition traffic flow rate and transition traffic flow rate simulation result of SCTM with FD and SLDS with SFD.

The log posterior of A_k is

$$\begin{aligned} \log p(A_k | D_k, \Sigma^k) &\propto \log p(D_k | A_k, \Sigma^k) p(A_k) \\ &= C - \frac{1}{2} \left(\sum_i (\mathbf{Q}_k^i - \bar{\mathbf{P}}_k^i \text{vec} \bar{A}_k)^T \right. \\ &\quad \cdot \Sigma^{-k} (\mathbf{Q}_k^i - \bar{\mathbf{P}}_k^i \text{vec} \bar{A}_k) \left. \right) \\ &\quad - \frac{1}{2} \left(\sum_i \text{vec} \bar{A}_k^T \Sigma_0^{-k} \text{vec} \bar{A}_k \right) \end{aligned} \quad (\text{A.7})$$

which can be rewritten as

$$\begin{aligned} \log p(A_k | D_k, \Sigma^k) &\propto C \\ &- \frac{1}{2} \left(\text{vec} \bar{A}_k^T \left(\Sigma_0^{-k} + \sum_i (\bar{\mathbf{P}}_k^i)^T \Sigma^{-k} (\bar{\mathbf{P}}_k^i) \right) \text{vec} \bar{A}_k \right. \\ &\quad \left. + \text{vec} \bar{A}_k^T \left(\sum_i (\bar{\mathbf{P}}_k^i)^T \Sigma^{-k} \mathbf{Q}_k^i \right) \right). \end{aligned} \quad (\text{A.8})$$

Conditioning on the data, we arrive at the desired posterior distribution:

$$p(A_k | D_k, \Sigma^k) = \mathcal{N}(\omega_j^k; \mathbf{N}, \Lambda) \quad (\text{A.9})$$

$$\Lambda = \left(\Sigma_0^{-k} + \sum_i (\bar{\mathbf{P}}_k^i)^T \Sigma^{-k} (\bar{\mathbf{P}}_k^i) \right)^{-1} \quad (\text{A.10})$$

$$\begin{aligned} \mathbf{N} &= \left(\Sigma_0^{-k} + \sum_i (\bar{\mathbf{P}}_k^i)^T \Sigma^{-k} (\bar{\mathbf{P}}_k^i) \right)^{-1} \\ &\quad \cdot \left(\sum_i (\bar{\mathbf{P}}_k^i)^T \Sigma^{-k} \mathbf{Q}_k^i \right). \end{aligned} \quad (\text{A.11})$$

We place an inverse-Wishart prior $IW(\Sigma^k; n_0, S_0)$; let $N_k = \{t | z_t = k, t = 1, 2, \dots, T\}$:

$$\begin{aligned} p(\Sigma^k) &= \frac{|S_0|^{n_0/2} |\Sigma^k|^{-(d+n_0+1)/2}}{2^{dn_0/2} \Gamma_d(n_0/2)} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^k S_0)\right), \end{aligned} \quad (\text{A.12})$$

where $\Gamma_d(\bullet)$ is the multivariate gamma function.

Put (A.12) and (A.6) into the following:

$$p(\Sigma^k | D, \omega^k) \propto p(D^k | \omega^k, \Sigma^k) p(\Sigma^k). \quad (\text{A.13})$$

The posterior of Σ^k is

$$p(\Sigma^k | D, \omega^k) = IW(n_0 + N_k, S_0 + S_{D_k}^k), \quad (\text{A.14})$$

where

$$S_{D_k}^k = \sum_i (\mathbf{Q}_k^i - \bar{\mathbf{P}}_k^i \text{vec} \bar{A}_k) (\mathbf{Q}_k^i - \bar{\mathbf{P}}_k^i \text{vec} \bar{A}_k)^T, \quad (\text{A.15})$$

Each precision parameter α_n^k is given a Gamma prior, $\mathcal{G}(a_0, b_0)$. The posterior of α_n^k is given by [14]

$$p(\alpha^k | A^k) = \mathcal{G}(a_N, b_N) \quad (\text{A.16})$$

$$\begin{aligned} a_N &= a_0 + \frac{p}{2} \\ b_N &= b_0 + \frac{\sum \omega_{i,j}^2}{2}. \end{aligned} \quad (\text{A.17})$$

B. The Proof of Remark 2

The log likelihood model in (17) is expanded as

$$\begin{aligned} \log p(D_k | A_k, \Sigma^k) &\propto C \\ &- \frac{1}{2} \left(\text{vec} \bar{A}_k^T \left(\sum_i (\bar{\mathbf{P}}_k^i)^T \Sigma^{-k} (\bar{\mathbf{P}}_k^i) \right) \text{vec} \bar{A}_k \right. \\ &\quad \left. + \text{vec} \bar{A}_k^T \left(\sum_i (\bar{\mathbf{P}}_k^i)^T \Sigma^{-k} \mathbf{Q}_k^i \right) \right. \\ &\quad \left. - \frac{1}{2} (\mathbf{Q}_k^i)^T \Sigma^{-k} (\mathbf{Q}_k^i) \right). \end{aligned} \quad (\text{B.1})$$

So $p(D_k | A_k, \Sigma^k)$ is

$$p(D_k | A_k, \Sigma^k) = C \exp(\Xi_1) \exp(\Xi_2), \quad (\text{B.2})$$

where

$$\Xi_1 = -\frac{1}{2} (\mathbf{Q}_k^i)^T \Sigma^{-k} (\mathbf{Q}_k^i) \quad (\text{B.3})$$

$$\begin{aligned} \Xi_2 &= -\frac{1}{2} \left(\text{vec} \bar{A}_k^T \left(\sum_i (\bar{\mathbf{P}}_k^i)^T \Sigma^{-k} (\bar{\mathbf{P}}_k^i) \right) \text{vec} \bar{A}_k \right. \\ &\quad \left. + \text{vec} \bar{A}_k^T \left(\sum_i (\bar{\mathbf{P}}_k^i)^T \Sigma^{-k} \mathbf{Q}_k^i \right) \right). \end{aligned} \quad (\text{B.4})$$

When $\exp(\Xi_1)$, $\exp(\Xi_2)$ is considered as $f(x, y)$, $\exp(\phi(\theta)^T u(x, y))$ in **Condition (1)** [18], and $g(\theta) = 1$, $\phi(\theta) = \text{vec} \bar{A}_k$, the likelihood model in (17) satisfies **Condition (1)**. In **Condition (2)** [18], $p(A_k)$ is said to be conjugate to the likelihood $p(D_k | A_k, \Sigma^k)$. By Appendix A, we know that the posterior $p(A_k | D_k, \Sigma^k) \propto p(D_k | A_k, \Sigma^k) p(A_k)$, of which the parametric is of the same form as the prior $p(A_k)$. So **Condition (2)** is satisfied.

In the same process, the posterior $p(A_k | D_k, \Sigma^k) = \mathcal{N}(\omega_j^k; \mathbf{N}, \Lambda)$ is $q_\theta(\theta)$ in **Theorem 1. (a)**. Note that \mathbf{N}, Λ are the expected complete data sufficient statistics $\bar{u}(y_i)$ mentioned in **Theorem 1. (a)** [18].

Data Availability

The traffic flow data used to support the findings of this study are included within the article, a brief introduction of which can be found in Sections 5.1 and 5.2. The open data can be visualized and downloaded from a public website in [22].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The research is supported by the National Natural Science Foundation of China under Grant no. 61773293.

References

- [1] J. Laval, Z. He, and F. Castrillon, "Stochastic extension of Newell's three-detector method," *Transportation Research Record*, vol. 2315, pp. 73–80, 2012.
- [2] A. F. Siqueira, C. J. T. Peixoto, C. Wu, and W.-L. Qian, "Effect of stochastic transition in the fundamental diagram of traffic flow," *Transportation Research Part B: Methodological*, vol. 87, pp. 1–13, 2016.
- [3] A. Sumalee, R. X. Zhong, T. L. Pan, and W. Y. Szeto, "Stochastic cell transmission model (SCTM): a stochastic dynamic traffic model for traffic state surveillance and assignment," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 507–533, 2011.
- [4] D. Ngoduy, "Multiclass first-order traffic model using stochastic fundamental diagrams," *Transportmetrica*, vol. 7, no. 2, pp. 111–125, 2011.
- [5] J. Li, Q.-Y. Chen, H. Wang, and D. Ni, "Analysis of LWR model with fundamental diagram subject to uncertainties," *Transportmetrica*, vol. 8, no. 6, pp. 387–405, 2012.
- [6] X. Wu, H. X. Liu, and N. Geroliminis, "An empirical analysis on the arterial fundamental diagram," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 255–266, 2011.
- [7] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.
- [8] C. F. Daganzo, "The cell transmission model, part II: network traffic," *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. 79–93, 1995.
- [9] 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.
- [10] X. Sun, L. Muñoz, and R. Horowitz, "Highway Traffic State Estimation Using Improved Mixture Kalman Filters for Effective Ramp Metering Control," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, pp. 6333–6338, USA, December 2003.
- [11] R. X. Zhong, A. Sumalee, T. L. Pan, and W. H. K. Lam, "Stochastic cell transmission model for traffic network with demand and supply uncertainties," *Transportmetrica A: Transport Science*, vol. 9, no. 7, pp. 567–602, 2013.
- [12] Z. Ghahramani, "An introduction to hidden Markov models and Bayesian networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 15, no. 1, pp. 9–42, 2001.
- [13] X. Wu and H. X. Liu, "A shockwave profile model for traffic flow on congested urban arterials," *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1768–1786, 2011.
- [14] M. Beal, *Variational Algorithms for Approximate Bayesian Inference [Ph.D. thesis]*, University College London, London, UK, 2003.
- [15] J. Q. Fan and Q. W. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*, pp. 12–123, Springer, New York, NY, USA, 2003.
- [16] R. Zhong, C. Chen, A. H. F. Chow, T. Pan, F. Yuan, and Z. He, "Automatic calibration of fundamental diagram for first-order macroscopic freeway traffic models," *Journal of Advanced Transportation*, vol. 50, no. 3, pp. 363–385, 2016.
- [17] J. Fan, M. Farnen, and I. Gijbels, "Local maximum likelihood estimation and inference," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 60, no. 3, pp. 591–608, 1998.
- [18] Z. Ghahramani and M. J. Beal, "Propagation algorithms for variational Bayesian learning," *Advances in Neural Information Processing Systems*, vol. 13, pp. 507–513, 2001.
- [19] A. Gelman, J. B. Carlin, H. S. Stern et al., *Bayesian Data Analysis (Third Edition)*, Chapman & Hall, 3rd edition, 2012.
- [20] Z. Ghahramani and G. E. Hinton, "Variational learning for switching state-space models," *Neural Computation*, vol. 12, no. 4, pp. 831–864, 2000.
- [21] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer-Verlag, New York, NY, USA, 2005.
- [22] "US Department of Transportation, NGSIM – Next Generation SIMulation".
- [23] H. K. Lo, "A cell-based traffic control formulation: strategies and benefits of dynamic timing plans," *Transportation Science*, vol. 35, no. 2, pp. 148–164, 2001.
- [24] J. Thai and A. M. Bayen, "State estimation for polyhedral hybrid systems and applications to the Godunov scheme for highway traffic estimation," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 60, no. 2, pp. 311–326, 2015.

Research Article

The Impact of Aggressive Driving Behavior on Driver-Injury Severity at Highway-Rail Grade Crossings Accidents

Changxi Ma ¹, Wei Hao ², Wang Xiang,² and Wei Yan³

¹School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou, China

²School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha, China

³Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong

Correspondence should be addressed to Wei Hao; haowei@csust.edu.cn

Received 10 June 2018; Revised 16 September 2018; Accepted 1 October 2018; Published 22 October 2018

Guest Editor: Zhiyuan Liu

Copyright © 2018 Changxi Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The effect of aggressive driving behavior on driver's injury severity is analyzed by considering a comprehensive set of variables at highway-rail grade crossings in the US. In doing so, we are able to use a mixed logit modelling approach; the study explores the determinants of driver-injury severity with and without aggressive driving behaviors at highway-rail grade crossings. Significant differences exist between drivers' injury severity with and without aggressive driving behaviors at highway-rail grade crossings. The level of injury for younger male drivers increases a lot if they are with aggressive driving behavior. In addition, driving during peak-hour is found to be a statistically significant predictor of high level injury severity with aggressive driving behavior. Moreover, environmental factors are also found to be statistically significant. The increased level of injury severity accidents happened for drivers with aggressive driving behavior in the morning peak (6-9 am), and the probability of fatality increases in both snow and fog condition. Driving in open space area is also found to be a significant factor of high level injury severity with aggressive driving behaviors. Bad weather conditions are found to increase the probability of drivers' high level injury severity for drivers with aggressive driving behaviors.

1. Introduction

Traffic accidents are a very important safety issue in the United States. For example, six million accidents were reported to the police in 2016 which involved 37461 fatalities [1]. Aggressive driving behaviors have been identified as a significant factor in traffic accidents [2]. Haleem and Gan [3] presented that aggressive driver manoeuvres contributed to more severe injuries. Aggressive driving behaviors were found to be involved in more than 55% of all fatal accidents [4]. A critical area of traffic safety research involves highway-rail grade crossing accidents, as vehicle-train collisions are one of the most dangerous traffic accidents in terms of fatalities, injuries, and property damage due to the average 4,000 to 1 weight ratio of trains to motor vehicles [5]. There were 25,945 highway-rail crossing accidents in the United States recent ten years' data between 2002 and 2011 in the FRA (Federal Railroad Administration) database (Highway-Rail

Grade Crossing Accident/Incident & US DOT Crossing Inventory Form).

Shinar [6] defined aggressive driving as the operation of a motor vehicle in a manner that endangers or is likely to endanger persons or property. The classification of driving behavior is based on FRA's original variable named "motorist" which included varied actions: (a) drove around or through the gate; (b) stopped and then proceeded; (c) did not stop; (d) stopped on crossing; (e) other. In our research, the actions "drove around or through the gate", "did not stop", and "stopped on crossing" are selected as aggressive driving behaviors. "Stopped and then proceeded" is selected as appropriate driving behaviors. As shown in Figure 1 for highway-rail grade crossing accidents, drivers with injury or fatality at highway-rail crossing accidents occurred more frequently in cases of aggressive driving than in cases without aggressive driving.

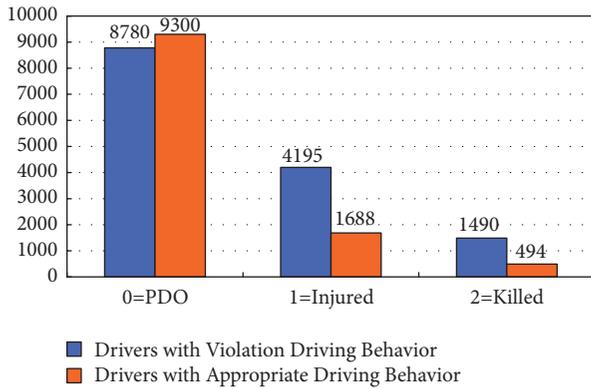


FIGURE 1: Highway-rail grade crossing accidents by driving behaviors.

Driver behavior at railway level crossings was modelled by Tey et al. [7] using mixed regression models. The analysis results show that active control devices produced much higher levels of driver compliance than passive control conditions. Another study by Tey et al. [8] revealed driver's responses at railway level crossings. This study describes and compares the driver response results from both field survey and driving simulator. Driving simulator is an important tool to analyze driving behaviors [9]. The conclusions show that different types of warning systems produce driver's behaviors at highway-rail grade crossings. According to previous research by Vanlaar et al. [10], there is a wide variety of behaviors which can be considered as aggressive driving behaviors. The data in this study was collected in a representative poll of 1,201 Canadian motor vehicle drivers. The results from a logistic regression model indicate that young male drivers are more likely to exhibit aggressive driving behaviors and have a history of traffic violation tickets. In addition, Hamdar et al. [11] provided an aggressive driving propensity calculation index at signalized intersections using structural equation modelling and applied this index to major signalized intersections in Washington DC. Millar [12] considered the influence of public self-consciousness and anger on aggressive driving. Kaysi and Abbany [13] developed binary probit models to analyze driving behavior on minor streets at unsignalized intersections. The most important variables for determining driving behavior were found to be driver age, car performance, and average vehicle speed on the major roads. Shinar and Compton [14] conducted a comprehensive study analyzing the relationship between aggressive driving behaviors with vehicle and driver's characteristics, as well as situational variables. The study collected over 2,000 occurrences of aggressive driving violations out of approximately 7,200 driver observations.

The level of injury study at highway-rail grade crossings has received the interest and attention from researchers [15]. Hao et al. [16–21] developed injury severity models to study the level of injury in different situations (control device types, area types, driver's age, gender, etc.). Guadamuz-Flores and Aguero-Valverde [22] used Full Bayesian Poisson-lognormal approaches to compare the effects of

various models, including heterogeneity-only, spatial-only, and heterogeneity-spatial models. The comparison results suggested that spatial correlation at highway-railway crossings should be considered in modelling of crash frequencies. Ghomi et al. [23] applied an ordered probit model, association rules, and classification and regression tree (CART) algorithms to the US Federal Railroad Administration's (FRA) HRGC accident database for the period 2007–2013 to identify VRU injury severity factors at HRGCs. Using six years of nationwide crashes from 2009 to 2014 in the US, Haleem [24] applied both the mixed logit and binary logit models based on the multiple predictors investigation (e.g., temporal crash characteristics, geometry, railroad, traffic, vehicle, and environment). The mixed logit model was found to outperform the binary logit model.

Eluru et al. [25] developed a latent class model to identify factors influencing driver-injury severity in highway-railway crossing accidents. The dataset used is from FRA (Federal Railroad Administration) highway-rail grade crossing data for 14,532 crossings from 1997 to 2006. Miranda-Moreno et al. [26] modelled and estimated each accident victims' injury severity at highway-rail grade crossings using multinomial models. A sample of 1773 crossings in Canada is considered in this study covering period from 1997 to 2004. Hu et al. [27] used a logit model to study important factors for injury severity at highway-rail crossings based on highway-grade crossing collision data from 1995 to 1997 in Taiwan. Zhang et al. [28] used a mixed logit model to conduct the analysis of drivers' route choice. McCollister et al. [29] developed an injury severity model to predict the probability of injury level at highway-rail crossings using FRA data. A logistic regression model was utilized as the methodology to estimate the probability of a fatality at highway-rail grade crossings.

Based on our literature review conducted, few previous studies have been found that investigated the influence of aggressive driving behaviors on driver's injury severity in highway-rail grade crossing accidents. This study aims to estimate the effect of aggressive driving behaviors on driver's injury severity in highway-rail grade crossing accidents in the United States using data from the Federal Railroad Administration (FRA) database covering the recent ten years' data since 2005. Using a mixed logit modelling approach, the study explores the determinants of driver-injury severity both with and without aggressive driving behaviors at highway-rail grade crossings.

2. Method

The methodology used in this research is to develop a mixed logit model to predict the probability of the level of injury in highway-rail grade crossing accidents based on data with three distinct driver-injury severity outcomes of property damage only, injury, and fatality. The traditional ordered models may not be suitable to consider level of injury severity due to restricting considering the influence of explanatory variables on severity outcomes [30]. Another possible consideration is to use random parameter ordered models; however there still exists the aforementioned limitation. The ordered logit and probit models are constrained to find only one

coefficient on each variable and it is in one direction, either towards higher severity or towards lower severity. It is a constraint because it is not inconceivable that a variable can increase both the probability of low and high severity. Also, a variable can tend towards middle severities and away from the low and high severities. However, the mixed logit model can capture heterogeneity through the use of random parameters. Not only that, but the mixed logit model allows explanatory variables to affect the mean of the distribution of the random parameters [31]. As a consequence, it is more reliable to adopt the unordered discrete outcome model to consider level of injury severity. Therefore, mixed logit model is utilized in this research and the following subsections describe the mixed logit model approach, the calculation of elasticity, and the likelihood test.

2.1. Mixed Logit Model. In this study, a mixed logit model (MXL) is used to predict the probability of the three driver-injury severity outcomes at highway-rail grade crossing accidents reported by police to estimate discrete driver-injury severity outcomes [32, 33],

$$S_{in} = \beta_i X_{in} + \varepsilon_{in} \quad (1)$$

where S_{in} is a severity function determining the driver-injury severity category i in highway-rail grade crossing accident n , X_{in} is a vector of independent variables which affect driver-injury severity category i in highway-rail grade crossing accident n , β_i is a vector of estimate parameters for driver-injury severity category i , and ε_{in} is an error term which is assumed to be generalized extreme value distributed. To arrive at the MXLs, random parameters are introduced with (β_i/φ) , where φ is a vector of variables of the chosen density function (mean and variance). The resulting mixed logit injury severity probabilities are

$$P_n \left(\frac{i}{\varphi} \right) = \int \frac{e^{\beta_i X_{in}}}{\sum_{\forall I} e^{\beta_i X_{in}}} \quad (2)$$

where $P_n(i/\varphi)$ is the probability of injury severity i conditional on $f(\beta_i/\varphi)$. If the variance in φ is determined to be explicitly different from zero, there will be incident-specific variations of the effect of X on injury severity across each crash observation n , with the density function $f(\beta_i/\varphi)$ used to determine the values of across crashes [34].

Due to the required numerical integration of the logit formula over the random, unobserved parameters, the maximum likelihood estimate of MXL is computationally complex. Thus, estimation methods based on simulation have been used. Research studies by McFadden and Ruud [35] and Geweke et al. [36] provide an in-depth development of simulation-based maximum likelihood methods for estimating MXL.

To evaluate the effect of individual variable estimates on injury severity outcome probabilities, an elasticity can be computed from the partial derivative for each observation n (n subscripting omitted) as

$$E_{X_{ki}}^{P(i/\varphi)} = \frac{\partial P(i/\varphi)}{\partial X_{ki}} \times \frac{X_{ki}}{P(i/\varphi)} \quad (3)$$

where $P(i/\varphi)$ is the probability of injury severity outcome i and X_k is the value of variable K . Elasticity values can be roughly interpreted as the percent effect that a 1% change in X_{ki} has on the injury severity outcome probability $P(i/\varphi)$. A pseudoelasticity indicator variable can be calculated to give the percent effect on the injury severity outcome probability of the parameter ranging in value from zero to one.

2.2. Likelihood Ratio Test. To determine whether there are significant differences between parameter estimates for drivers with and without aggressive driving behaviors, a likelihood ratio test was performed [37–40]. The ratio test determines the transferability of aggressive driving behavior model's coefficients developed in the aggressive driving behavior model to the appropriate driving behavior model and whether there is a significant difference between the two groups. The group of the variables is used to develop a joint model for both conditions. The likelihood ratio statistic is

$$LR = -2(L_j - L_L - L_U) \quad (4)$$

where $J = K_L + K_U - K_j$, K_j , K_L , and K_U are the number of coefficients in the joint model, the aggressive driving behavior model, and the appropriate driving behavior model, respectively, and L_j , L_L , and L_U represent the log-likelihoods at convergence for the joint model, the aggressive driving behavior model, and the appropriate driving behavior model, respectively.

The null hypothesis for (4) is that the restricted model (joint model) does not have a lower log-likelihood compared with the unrestricted models (separate aggressive driving and appropriate driving behavior models), indicating a lack of significant difference between the driving behavior-specific models and the joint model.

The transferability of coefficients from the aggressive driving behavior model to the appropriate driving behavior model was also investigated. For instance, a model is estimated for the appropriate driving behavior. The resulting model is applied to data for the behaviors during aggressive driving situations with all coefficients restricted to the values estimated for the situations without aggressive driving behaviors, yielding the restricted log-likelihood. Then the exact same model specification is set free and coefficients estimated on the data for the aggressive driving condition, yielding the unrestricted log-likelihood L_R . The likelihood ratio statistic is then

$$LR = -2(L_R - L_U) \quad (5)$$

where LR is χ^2 distributed with degrees of freedom J , where J is the number of restrictions.

The null hypothesis for (5) is that the coefficients are equal for both the aggressive driving and appropriate driving behavior models. If transferability is rejected at a high significance (the significance level is at 0.05), it is statistically determined that the aggressive driving model and appropriate driving behavior model are not equivalent and behaviors with aggressive driving behaviors have an impact on crash severity.

TABLE 1: Description of Highway-Rail Collision Characteristics.

Description		Drivers with Appropriate Driving Behaviors		Drivers with Aggressive Driving Behaviors	
		Frequency	Percentage	Frequency	Percentage
Dependent Variable					
Driver	0= PDO	9300	81.00%	8780	60.70%
	1= injured	1688	14.70%	4195	29.00%
	2= killed	492	4.30%	1490	10.30%
Independent Variables					
Vehicle Speed	0 (Less than 50mph)	11470	99.90%	13871	95.90%
	1 (more than 50mph)	11	0.10%	593	4.10%
Train Speed	0 (Less than 50mph)	9587	83.50%	12540	86.70%
	1 (more than 50mph)	1894	16.50%	1924	13.30%
Age	0 (More than 25)	10563	92.00%	13235	91.50%
	1 (Less than 25)	953	8.00%	1085	7.50%
Gender	0 (Female)	3146	27.40%	3124	21.60%
	1 (Male)	8335	72.60%	11340	78.40%
Control Device	0(Active Control)	7659	66.7%	8144	56.31%
	1(Passive Control)	3822	33.3%	6320	43.69%
Vehicle Type	1(Sedan)	5235	45.60%	6581	45.50%
	2(Truck)	758	6.60%	1287	8.90%
	3(SUV)	1803	15.70%	2184	15.10%
	4(Pick-up)	1447	12.60%	2705	18.70%
	5(Van)	402	3.50%	663	4.60%
	6(Bus)	34	0.30%	14	0.10%
	7(Other)	1791	15.60%	1041	7.20%
Time Period	0 (Other Time Periods)	8209	71.50%	9893	68.40%
	1(Peak hour including 7:00AM-9:00AM & 4:00PM-6:00PM)	3272	28.50%	4571	31.60%
Weather	1 (Cloudy)	2296	20.00%	2849	19.70%
	2 (Rain)	792	6.90%	897	6.20%
	3 (Fog)	172	1.50%	217	1.50%
	4 (Sleet)	34	0.30%	29	0.20%
	5 (Snow)	321	2.80%	347	2.40%
	6 (Clear)	7864	68.50%	10125	70%
Accident Happened Area	0(other areas)	9403	81.90%	11253	77.80%
	1(open space)	2078	18.10%	3211	22.20%

3. Data

A complete and detailed data collection is essential to make sure to get reliable conclusions. The original dataset obtained from the FRA database is made up of 25,945 highway-rail grade crossing accidents which occurred in the United States recent ten years' data since 2005. Injury severity is the dependent variable in this study which is ranked as 0-property damage only (PDO), 1-injury, and 2-fatality. For the aggressive driving behavior dataset, the distribution of crashes by injury level is as follows: 60.70% PDO, 29.00% injured, and 10.30% fatality. For the appropriate driving behavior group, the distribution of crashes by injury level is

as follows: 81.00% PDO, 14.70% injury, and 4.30% fatality. The independent variables in this study include schedule factor, vehicle speed, vehicle type, weather condition, train speed, driver's age and gender, control device type, and area type. Table 1 shows the frequency and percentage distribution of these variables.

4. Model Results and Discussion

Two MXL models are estimated based on driving behavior and a likelihood ratio test was utilized to test if significant differences existed between parameter estimates for these two models. These tests indicate that the hypothesis that

aggressive driving behavior model and appropriate driving behavior model are the same can be rejected at a confidence level exceeding 99.50%. The likelihood ratio tests included a comparison of a combined aggressive driving behavior model/appropriate driving behavior model with separate aggressive driving behavior and appropriate driving behavior models (see (4)) and a comparison of aggressive driving -converged coefficients based on appropriate driving behavior data with appropriate driving behavior-converged coefficients based on aggressive driving behavior data (see (5)). In light of these results, separate models for aggressive driving behavior model and appropriate driving behavior model are developed to capture the differences.

Tables 2 and 3 show the results of the mixed logit estimation. The models included all estimated parameters that are statistically significant. Parameters which produced statistically significant standard errors for their assumed distribution were found to be random and the standard errors for these parameters are shown in the tables. The parameters for which estimated standard errors were not statistically different from 0 were found to be fixed and these parameters are denoted as fixed in the tables. For all of the random parameters, the normal distribution was found to provide the best statistical fit.

4.1. Driver's Characteristics. Looking at the effect of driver age, the findings on coefficient estimation results shown in Tables 2 and 3 indicate that younger drivers are more likely to exhibit aggressive driving behavior with a 15.2% increase in the probability of a fatality in Table 2 given that a highway-rail grade crossing accident happened compared to a 6.9% increase for drivers with appropriate driving behaviors in Table 3. Several previous studies provide support for this finding. One reason may be that younger drivers (less than 25) fundamentally underestimate the risk of being involved in a crash [41]. Another study found that there is a substantial difference in driving behaviors across the different age categories [14]. The probability of injury severity for younger drivers with aggressive driving behaviors is extremely higher than older drivers.

Turning to the influence of gender, this study found that male drivers were more likely to express aggressive driving behaviors outwardly with high level injury severity given a highway-rail grade crossing accident happened than females. A previous study found significant physiological differences between the genders and males who endorse an exaggerated male stereotype are more likely to engage in aggressive driving behaviors [42]. Women have a stronger sense of obligation to traffic laws and tend to obey those laws while men tend to overestimate their driving ability and underestimate the risks associated with traffic violations [43–45].

4.2. Vehicle and Train Characteristics. The probability of a fatality increases by 14.8% for drivers with aggressive driving behaviors and by 8.4% for drivers without aggressive driving behaviors when the vehicle driving speed exceeded 50 mph. Consistent with previous studies, it is expected that, as the speeds of vehicles at railroad crossings increase, the injury

severity levels in accidents also increase [16]. The impact of vehicle speeds on injury severity may be explained by the fact that higher vehicle speeds will result in the inability of drivers to visually detect an on-coming train, thereby increasing the likelihood of a higher injury severity in the event of a collision [17]. As a result, we can speculate that driving at high speed together with aggressive driving behaviors such as “drove through the gate” and “did not stop” at highway-rail grade crossings is really dangerous for vehicle drivers.

Meanwhile, the injury severity model results in Tables 2 and 3 show that higher train speed increases the probability of a fatality by 7.8% for drivers with aggressive driving behaviors and 3.8% for drivers with appropriate driving behaviors at highway-rail grade crossings. As a consequence, train speed limit reductions could help to moderate injury severity by allowing for more time for last minute maneuvering and braking actions to avoid collisions or lessen their severity [19].

Considering the impact of vehicle type, it is found that SUV and pick-up drivers are more likely to have a higher injury severity when they drive with aggressive violation behaviors. The model results in Table 2 show that the probability of fatality increases by 8.9% for SUV drivers with aggressive driving behaviors and by 8.1% for pick-up drivers with aggressive driving behaviors. This finding is confirmed by a previous study [5], which found that drivers who drive an SUV or a pick-up truck are more likely to be severely injured in accidents during rush hour.

4.3. Environmental/Situational Factors. The increased level of injury severity accidents happened for drivers with aggressive driving behavior in the morning peak (6–9 am). It could be explained as a reflection of the time pressures on drivers to reach their place of employment on time. In highway-rail grade crossing accidents, the probability of fatality during peak hours increased by 15.1% for drivers with aggressive driving behaviors in Table 2 and by 3.5% for drivers with appropriate driving behaviors in Table 3. As indicated by another study, time pressures, when combined with traffic congestion, can cause driver's aggressive driving behaviors [46, 47]. Both time pressures and traffic congestion are common during the peak-hour period which is consistent with previous studies [48].

For weather conditions, the model results in Table 2 show that the probability of fatality increases by 2.5% (snow condition) and 5.5% (fog condition) with driving aggressive driving behavior relative to clear weather condition. Previous research has shown that bad weather makes roads less skid resistant and decreases visibility which results in poorer braking and steering performance and worse impact angles leading to more severe injuries [19]. The results in this study show that foggy conditions particularly increase the probability of a fatality occurring in highway-rail grade crossing accidents for drivers with aggressive driving behaviors as a result of poor visibility.

4.4. Highway-Rail Grade Crossing Attributes. Open space area is found to increase the probability of fatal accidents by 12.3% in Table 2 for drivers with aggressive driving behaviors and by 9.1% in Table 3 for drivers with appropriate driving

TABLE 2: Mixed Logit Model Estimation Results for Drivers with Aggressive Driving behavior.

Variables Description	Coefficient	Standard Error	t-Statistic	Elasticity		
				Fatality	Injury	PDO
Defined for Fatality						
Constant	5.525(1.632)	0.682(0.371)	13.16(3.56)			
<i>Vehicle & Train Characteristics</i>						
Vehicle Speed>50mph	7.875(0.198)	0.092(0.127)	7.51(2.74)	14.8%	13.2%	-2.8%
Train Speed>50mph	4.383(0.231)	0.127(0.235)	5.63(1.59)	7.8%	4.8%	-5.5%
SUV	2.351(0.052)	0.096(0.192)	5.23(1.86)	8.9%	-4.5%	3.8%
Pick-up	1.529(0.053)	0.321(0.429)	3.15(1.96)	8.1%	-25.2%	2.5%
<i>Driver's Characteristics</i>						
Age less than 25	2.351(1.321)	0.563(0.375)	7.51(2.54)	15.2%	17.3%	-3.2%
Male, fixed parameter	1.127		5.23	4.4%	5.8%	-2.9%
<i>Environmental/ situational factors</i>						
Peak-hour, fixed parameter	2.641	0.257	5.21	15.1%	16.2%	-5.5%
Weather fog	1.235(0.276)	0.087(0.382)	-3.81(3.91)	5.5%	-4.3%	1.8%
Weather Snow	1.682(0.357)	0.537(0.291)	-5.52(1.69)	2.5%	-1.8%	2.3%
<i>Highway-rail Grade Crossing Attributes</i>						
Passive control, fixed parameter	2.837	0.892	6.83	7.9%	5.9%	-4.3%
Open space area	2.582(0.179)	0.097(0.476)	4.35(3.94)	12.3%	7.6%	-8.3%
Defined for Injury						
<i>Vehicle & Train Characteristics</i>						
Vehicle Speed>50mph	5.252(0.232)	0.137(0.215)	4.96(2.58)	13.7%	11.9%	-3.7%
Train Speed>50mph	4.327(0.384)	0.261(0.198)	5.25(2.63)	6.9%	5.3%	-6.2%
Pick-up	1.286(0.097)	0.375(0.413)	4.86(4.05)	7.2%	-13.2%	3.8%
<i>Driver's Characteristics</i>						
Age less than 25	1.232(1.273)	0.678(0.421)	6.51(1.24)	13.3%	12.7%	-4.2%
<i>Environmental/ situational factors</i>						
Peak-hour, fixed parameter	2.321		4.52	7.5%	5.3%	-8.2%
<i>Highway-rail Grade Crossing Attributes</i>						
Passive control, fixed parameter	2.257		5.57	3.5%	4.2%	-2.5%
Defined for PDO						
<i>Vehicle & Train Characteristics</i>						
Train Speed>50mph	4.751(0.563)	0.317(0.385)	6.69(2.31)	7.8%	4.1%	-5.3%
SUV	2.751(0.127)	0.291(0.202)	4.96(3.53)	5.3%	2.7%	-3.6%
<i>Driver's Characteristics</i>						
Age less than 25	1.263(2.081)	0.893(0.563)	4.59(5.87)	10.5%	9.8%	-8.5%
Male, fixed parameter	0.761		6.17	3.9%	5.2%	-1.9%
<i>Environmental/ situational factors</i>						
Weather fog	1.382(0.344)	0.256(0.567)	-5.92(5.76)	4.8%	3.9%	-7.1%
Weather Snow	1.392(0.467)	0.763(0.361)	-4.93(4.51)	3.6%	-1.7%	3.2%
<i>Highway-rail Grade Crossing Attributes</i>						
Passive control, fixed parameter	2.027		5.18	2.9%	3.1%	-5.1%
Open space area	1.121(0.532)	0.345(1.321)	7.53(4.53)	10.5%	6.5%	-7.8%
<i>Model Statistics</i>						
Number of Observations		14464				
Log-likelihood at constants		-1586.17				
Log-likelihood at convergence		-863.25				

Notes: Parentheses indicate standard errors of random parameter estimates.

TABLE 3: Mixed Logit Model Estimation Results for Drivers with Appropriate Driving Behavior.

Variables Description	Parameter Estimate	Standard Error	t-Statistic	Elasticity		
				Fatality	Injury	PDO
Defined for Fatality						
Constant	-4.761(1.572)	0.1683(0.523)	8.75(2.38)			
<i>Vehicle & Train Characteristics</i>						
Vehicle Speed	3.386(0.015)	0.007(0.012)	3.28(1.56)	8.4%	10.3%	-5.2%
Train Speed	2.563(0.036)	0.062(0.032)	5.62(2.91)	3.8%	3.2%	-2.3%
SUV	1.753(0.028)	0.126(0.328)	4.82(1.78)	7.2%	-8.3%	5.1%
Pick-up	0.952(0.041)	0.203(0.512)	3.38(1.28)	3.5%	-31.2%	1.9%
<i>Driver's Characteristics</i>						
Age less than 25	1.361(0.008)	0.006(0.026)	6.32(3.52)	6.9%	12.1%	-6.5%
Male, fixed parameter	0.525	0.128	3.82	2.1%	3.6%	-1.7%
<i>Environmental/ situational factors</i>						
Peak-hour, fixed parameter	1.952	0.628	4.02	3.5%	12.3%	-3.7%
Weather fog	0.652(0.192)	0.072(0.031)	5.12(2.37)	3.9%	-5.6%	2.3%
Weather Rain	0.851(0.325)	0.627(0.043)	2.85(1.21)	1.2%	-2.6%	0.8%
<i>Highway-rail Grade Crossing Attributes</i>						
Passive control, fixed parameter	1.572	0.521	6.57	5.2%	3.6%	-2.7%
Open space area	1.378(0.258)	0.063(0.321)	3.31(2.19)	9.1%	5.2%	-3.6%
Defined for Injury						
<i>Vehicle & Train Characteristics</i>						
Vehicle Speed	2.321(0.021)	0.015(0.009)	3.12(2.16)	10.7%	8.4%	-1.9%
SUV	1.521(0.032)	0.071(0.027)	3.61(1.37)	6.5%	-14.9%	1.7%
<i>Driver's Characteristics</i>						
Male, fixed parameter	0.536	0.218	5.32	1.9%	2.5%	-0.9%
<i>Environmental/ situational factors</i>						
Peak-hour, fixed parameter	1.582	0.372	3.86	5.8%	4.2%	-9.1%
<i>Highway-rail Grade Crossing Attributes</i>						
Open space area	1.385(0.327)	0.051(0.273)	5.52(1.39)	7.3%	3.6%	-1.9%
Defined for PDO						
<i>Vehicle & Train Characteristics</i>						
Vehicle Speed	1.251(0.032)	0.029(0.005)	5.12(1.58)	6.2%	4.9%	-5.1%
Train Speed	1.385(0.027)	0.031(0.038)	6.23(2.16)	3.2%	2.7%	-3.2%
Pick-up	0.891(0.018)	0.056(0.043)	3.76(2.32)	2.1%	-17.8%	2.6%
<i>Driver's Characteristics</i>						
Age less than 25	0.683(0.007)	0.012(0.021)	4.58(3.09)	6.9%	5.6%	-4.7%
Male, fixed parameter	0.352	0.235	3.98	1.7%	2.3%	-2.6%
<i>Environmental/ situational factors</i>						
Peak-hour, fixed parameter	1.305	0.528	4.31	3.7%	2.5%	-7.3%
Weather Rain	0.792(0.217)	0.479(0.218)	0.68(2.51)	2.9%	2.5%	-4.6%
Weather Sleet	0.631(0.129)	0.762(0.327)	5.38(3.62)	1.9%	-0.7%	2.5%
<i>Highway-rail Grade Crossing Attributes</i>						
Passive control, fixed parameter	1.127	0.436	4.58	2.3%	1.8%	-3.1%
<i>Model Statistics</i>						
Number of Observations		11481				
Log-likelihood at constants		-1032.07				
Log-likelihood at convergence		-576.18				

Notes: Parentheses indicate standard errors of random parameter estimates.

behaviors. This can be explained by the fact that drivers may drive more recklessly with aggressive driving behaviors in open space areas compared to high population density areas [49]. In addition, an explanation for this interesting result is that operating speeds in an open space area may tend to be higher than in other areas as open space areas generally have lower traffic densities which increases the likelihood of high level injury severity given an accident happened [48].

For control device types, consistent with previous studies [18, 50], passive control was found positively correlated with high level injury severity for drivers with aggressive driving behaviors which is demonstrated by our research. Passive control devices are found to increase the probability of a fatal accident by 7.9% in Table 2 for drivers with aggressive driving behaviors and by 5.2% in Table 3 for drivers with appropriate driving behaviors.

4.5. Suggestions. According to the above result, many factors could influence aggressive driving behaviors and injury severity. The comprehensive countermeasures should be used to limit the aggressive driving behaviors. Firstly, speed limit measures are necessary to reduce vehicles' speed (less than 50mph) when approaching highway-rail grade crossing, such as the posted speed limit, yield sign, stop sign, and stop line sign. And it is better to also reduce the train's speed when approaching HRGC. Secondly, the warning devices are important to remind drivers keeping vigilant, such as flashing lights, pavement markings, and photo/audio/video enforcement, especially in the morning peak (6-9 am) and adverse weather (fog and snow). Finally, improving the HRGC condition is also important, such as traffic signal, greater sight distance, and grade separation/closure. In addition, for highway-rail grade crossing accident improvement measure, Saccomanno et al. [51] proposed a Bayesian data fusion method could estimate countermeasure effects for reducing collisions at highway-railway grade crossings.

5. Conclusions

This paper explores the injury severity of drivers with aggressive driving behaviors utilizing accident data at US highway-rail grade crossings. The results of this study have implications for those who aim to reduce accidents involving aggressive driving behaviors. The estimation results indicate that driver age and gender as well as environmental, time-of-day, and weather characteristics have a strong effect on the injury severity level in highway-rail grade crossing accidents. Younger drivers, particularly young males, who engage in aggressive driving behavior during peak hours are more likely to suffer more severe injuries. As a result, education and enforcement efforts aimed at younger male drivers could significantly improve safety by reducing aggressive driving behaviors. In addition, accidents in open space areas are more likely to involve aggressive driving behaviors and produce severe driver injuries. Bad weather and peak-hour conditions also increase the likelihood that drivers will experience severe injuries in accidents when they engage in aggressive driving behaviors in these situations. Injuries also tended to be more severe in open space areas with passive control for drivers

with aggressive driving behaviors. As a result, the findings offer insights into the effects of drivers with aggressive driving behavior on driver's injury severity at highway-rail grade crossing accidents.

Data Availability

The safety data used to support the findings of this study were supplied by FRA (Federal Railroad Administration) and you can access the data by the following website: <https://safetydata.fra.dot.gov/OfficeofSafety/default.aspx>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by National Natural Science Foundation of China (nos. 71861023 and 71701023), Program of Humanities and Social Science of Education Ministry of China (no. 18YJC630118), Lanzhou Jiaotong University (no. 201804), Hunan Key Laboratory of Smart Roadway and Cooperative Vehicle-Infrastructure Systems (no. 2017TP1016), and the Outstanding Youth Foundation of Hunan Education Department (no. 17B015). The authors express their thanks to all who participated in this research for their cooperation.

References

- [1] N. Kalra and D. Groves, *The Enemy of Good: Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles*, RAND Corporation, 2017.
- [2] R. Paleti, N. Eluru, and C. R. Bhat, "Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes," *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 1839–1854, 2010.
- [3] K. Haleem and A. Gan, "Contributing factors of crash injury severity at public highway-railroad grade crossings in the U.S.," *Journal of Safety Research*, vol. 53, pp. 23–29, 2015.
- [4] X. Du, Y. Shen, R. Chang, and J. Ma, "The exceptionists of Chinese roads: The effect of road situations and ethical positions on driver aggression," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 58, pp. 719–729, 2018.
- [5] X. Yan, L. D. Han, S. Richards, and H. Millegan, "Train-vehicle crash risk comparison between before and after stop signs installed at highway-rail grade crossings," *Traffic Injury Prevention*, vol. 11, no. 5, pp. 535–542, 2010.
- [6] D. Shinar, "Aggressive driving: The contribution of the drivers and the situation," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 1, no. 2, pp. 137–160, 1998.
- [7] L.-S. Tey, G. Wallis, S. Cloete, and L. Ferreira, "Modelling driver behaviour towards innovative warning devices at railway level crossings," *Accident Analysis & Prevention*, vol. 51, pp. 104–111, 2013.
- [8] L.-S. Tey, L. Ferreira, and A. Wallace, "Measuring driver responses at railway level crossings," *Accident Analysis & Prevention*, vol. 43, no. 6, pp. 2134–2141, 2011.
- [9] W. Yan, W. Xiang, S. C. Wong, X. Yan, Y. C. Li, and W. Hao, "Effects of hands-free cellular phone conversational cognitive

- tasks on driving stability based on driving simulation experiment," *Transportation Research Part F: Traffic Psychology & Behaviour*, no. 58, pp. 264–281, 2018.
- [10] W. Vanlaar, H. Simpson, D. Mayhew, and R. Robertson, "Aggressive driving: A survey of attitudes, opinions and behaviors," *Journal of Safety Research*, vol. 39, no. 4, pp. 375–381, 2008.
- [11] S. H. Hamdar, H. S. Mahmassani, and R. B. Chen, "Aggressiveness propensity index for driving behavior at signalized intersections," *Accident Analysis & Prevention*, vol. 40, no. 1, pp. 315–326, 2008.
- [12] M. Millar, "The influence of public self-consciousness and anger on aggressive driving," *Personality and Individual Differences*, vol. 43, no. 8, pp. 2116–2126, 2007.
- [13] I. A. Kaysi and A. S. Abbany, "Modeling aggressive driver behavior at unsignalized intersections," *Accident Analysis & Prevention*, vol. 39, no. 4, pp. 671–678, 2007.
- [14] D. Shinar and R. Compton, "Aggressive driving: an observational study of driver, vehicle, and situational variables," *Accident Analysis & Prevention*, vol. 36, no. 3, pp. 429–437, 2004.
- [15] M. Zhang, A. J. Khattak, J. Liu, and D. Clarke, "A comparative study of rail-pedestrian trespassing crash injury severity between highway-rail grade crossings and non-crossings," *Accident Analysis & Prevention*, no. 117, pp. 427–438, 2018.
- [16] W. Hao and J. Daniel, "Severity of injuries to motor vehicle drivers at highway-rail grade crossings in the United States," *Transportation Research Record*, no. 2384, pp. 102–108, 2013.
- [17] W. Hao and J. Daniel, "Motor vehicle driver injury severity study under various traffic control at highway-rail grade crossings in the United States," *Journal of Safety Research*, vol. 51, pp. 41–48, 2014.
- [18] W. Hao, C. Kamga, and J. Daniel, "The effect of age and gender on motor vehicle driver injury severity at highway-rail grade crossings in the United States," *Journal of Safety Research*, vol. 55, pp. 105–113, 2015.
- [19] W. Hao and J. Daniel, "Driver injury severity related to inclement weather at highway-rail grade crossings in the United States," *Traffic Injury Prevention*, vol. 17, no. 1, pp. 31–38, 2016.
- [20] W. Hao, C. Kamga, and D. Wan, "The effect of time of day on driver's injury severity at highway-rail grade crossings in the United States," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 3, no. 1, pp. 37–50, 2016.
- [21] W. Hao and C. Kamga, "The effects of lighting on driver's injury severity at highway-rail grade crossings," *Journal of Advanced Transportation*, vol. 50, no. 4, pp. 446–458, 2016.
- [22] R. Guadamuz-Flores and J. Aguero-Valverde, "Bayesian Spatial Models of Crash Frequency at Highway–Railway Crossings," *Transportation Research Record*, vol. 2608, no. 1, pp. 27–35, 2017.
- [23] H. Ghomi, M. Bagheri, L. Fu, and L. F. Miranda-Moreno, "Analyzing injury severity factors at highway railway grade crossing accidents involving vulnerable road users: A comparative study," *Traffic Injury Prevention*, vol. 17, no. 8, pp. 833–841, 2016.
- [24] K. Haleem, "Investigating risk factors of traffic casualties at private highway-railroad grade crossings in the United States," *Accident Analysis & Prevention*, vol. 95, pp. 274–283, 2016.
- [25] N. Eluru, M. Bagheri, L. F. Miranda-Moreno, and L. Fu, "A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings," *Accident Analysis & Prevention*, vol. 47, pp. 119–127, 2012.
- [26] L. F. Miranda-Moreno, L. Fu, S. Ukkusuri, and D. Lord, "How to incorporate accident severity and vehicle occupancy into the hot spot identification process?" *Transportation Research Record*, no. 2102, pp. 53–60, 2009.
- [27] S.-R. Hu, C.-S. Li, and C.-K. Lee, "Investigation of key factors for accident severity at railroad grade crossings by using a logit model," *Safety Science*, vol. 48, no. 2, pp. 186–194, 2010.
- [28] W. B. Zhang, Y. Qi, Y. Yan, J. J. Tang, and Y. H. Wang, "A method of and traveller behavior analysis under multimodal traffic condition," *Transportation Research Part D: Transport & Environment*, no. 52, pp. 139–155, 2017.
- [29] G. M. McCollister and C. C. Pflaum, "A model to predict the probability of highway rail crossing accidents," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 221, no. 3, pp. 321–329, 2007.
- [30] S. Yasmin and N. Eluru, "Evaluating alternate discrete outcome frameworks for modeling crash injury severity," *Accident Analysis & Prevention*, vol. 59, pp. 506–521, 2013.
- [31] J.-K. Kim, G. F. Ulfarsson, S. Kim, and V. N. Shankar, "Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender," *Accident Analysis & Prevention*, vol. 50, pp. 1073–1081, 2013.
- [32] Z. Feng, J. Zhan, C. Ma et al., "Is cognitive intervention or forgiveness intervention more effective for the reduction of driving anger in Chinese bus drivers?" *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 55, pp. 101–113, 2018.
- [33] W. Zhang, Z. Hu, Z. Feng, C. Ma, K. Wang, and X. Zhang, "Investigating factors influencing drivers' speed selection behavior under reduced visibility conditions," *Traffic Injury Prevention*, vol. 19, no. 5, pp. 488–494, 2018.
- [34] D. McFadden and K. Train, "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, vol. 15, no. 5, pp. 447–470, 2000.
- [35] D. McFadden and P. A. Ruud, "Estimation by simulation," *Review of Economics and Statistics*, vol. 76, no. 4, pp. 591–608, 1994.
- [36] J. Geweke, M. Keane, and D. Runkle, "Alternative computational approaches to inference in the multinomial probit model," *Review of Economics and Statistics*, vol. 76, no. 4, pp. 609–632, 1994.
- [37] G. F. Ulfarsson and F. L. Mannering, "Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents," *Accident Analysis & Prevention*, vol. 36, no. 2, pp. 135–147, 2004.
- [38] J. Tang, J. Liang, S. Zhang, H. Huang, and F. Liu, "Inferring driving trajectories based on probabilistic model from large scale taxi GPS data," *Physica A: Statistical Mechanics and its Applications*, vol. 506, pp. 566–577, 2018.
- [39] Y. Yan, S. Zhang, J. Tang, and X. Wang, "Understanding characteristics in multivariate traffic flow time series from complex network structure," *Physica A: Statistical Mechanics and its Applications*, vol. 477, pp. 149–160, 2017.
- [40] J. Tang, S. Zhang, X. Chen, F. Liu, and Y. Zou, "Taxi trips distribution modeling based on Entropy-Maximizing theory: A case study in Harbin city—China," *Physica A: Statistical Mechanics and its Applications*, vol. 493, pp. 430–443, 2018.
- [41] S. Islam and F. Mannering, "Driver aging and its effect on male and female single-vehicle accident injuries: Some additional evidence," *Journal of Safety Research*, vol. 37, no. 3, pp. 267–276, 2006.

- [42] R. Lawton and A. Nutter, "A comparison of reported levels and expression of anger in everyday and driving situations," *British Journal of Psychology*, vol. 93, no. 3, pp. 407–423, 2002.
- [43] B. Krahe and I. Fenske, "Predicting Aggressive Driving Behavior: The Role of Macho Personality, Age, and Power of Car," *Aggressive Behavior*, vol. 28, no. 1, pp. 21–29, 2002.
- [44] K. Jiang, F. Ling, Z. Feng et al., "Effects of mobile phone distraction on pedestrians' crossing behavior and visual attention allocation at a signalized intersection: An outdoor experimental study," *Accident Analysis & Prevention*, vol. 115, pp. 170–177, 2018.
- [45] N. Lyu, L. Xie, C. Wu, Q. Fu, and C. Deng, "Driver's cognitive workload and driving performance under traffic sign information exposure in complex environments: A case study of the highways in China," *International Journal of Environmental Research and Public Health*, vol. 14, no. 2, article no. 203, 2017.
- [46] D. Parker, T. Lajunen, and H. Summala, "Anger and aggression among drivers in three European countries," *Accident Analysis & Prevention*, vol. 34, no. 2, pp. 229–235, 2002.
- [47] L. Buckley and M. S. Foss, "Protective factors associated with young passenger intervening in risky driving situations," *Journal of Safety Research*, vol. 43, no. 5-6, pp. 351–356, 2012.
- [48] C. Ma, R. He, W. Zhang, and X. Ma, "Path optimization of taxi carpooling," *PLoS ONE*, vol. 13, no. 8, pp. 1–15, 2018.
- [49] K. H. Beck, M. Q. Wang, and M. M. Mitchell, "Concerns, dispositions and behaviors of aggressive drivers: What do self-identified aggressive drivers believe about traffic safety?" *Journal of Safety Research*, vol. 37, no. 2, pp. 159–165, 2006.
- [50] C. Ma, W. Hao, A. Wang, and H. Zhao, "Developing a Coordinated Signal Control System for Urban Ring Road under the Vehicle-infrastructure Connected Environment," *IEEE Access*, vol. 6, pp. 52471–52478, 2018.
- [51] F. F. Saccomanno, P. Y.-J. Park, and L. Fu, "Estimating countermeasure effects for reducing collisions at highway-railway grade crossings," *Accident Analysis & Prevention*, vol. 39, no. 2, pp. 406–416, 2007.

Research Article

Driving Risk Detection Model of Deceleration Zone in Expressway Based on Generalized Regression Neural Network

Weiwei Qi ^{1,2}, Zhexuan Wang,¹ Ruru Tang,³ and Linhong Wang ⁴

¹School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China

²School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia

³School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

⁴School of Transportation, Jilin University, Changchun 130022, China

Correspondence should be addressed to Linhong Wang; wanghonglin0520@126.com

Received 20 July 2018; Accepted 24 September 2018; Published 10 October 2018

Guest Editor: David Z. W. Wang

Copyright © 2018 Weiwei Qi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Drivers' mistakes may cause some traffic accidents, and such accidents can be avoided if prompt advice could be given to drivers. So, how to detect driving risk is the key factor. Firstly, the selected parameters of vehicle movement are reaction time, acceleration, initial speed, final speed, and velocity difference. The ANOVA results show that the velocity difference is not significant in different driving states, and the other four parameters can be used as input variables of neural network models in deceleration zone of expressway, which have fifteen different combinations. Then, the detection model results indicate that the prediction accuracy rate of testing set is up to 86.4%. An interesting finding is that the number of input variables is positively correlated with the prediction accuracy rate. By applying the method, the dangerous state of vehicles could be released through mobile internet as well as drivers' start of risky behaviors, such as fatigue driving, drunk driving, speeding driving, and distracted driving. Numerical analyses have been conducted to determine the conditions required for implementing this detection method. Furthermore, the empirical results of the present study have important implications for the reduction of crashes.

1. Introduction

The road traffic system consists of drivers, roads, vehicles, and environment. Traffic safety is an important part in the field of social management, and 90% of traffic accidents are caused by drivers' risky behavior [1, 2]. There are four typical risky driving behaviors: fatigue driving, drunk driving, speeding driving, and distracted driving. Human, vehicle, road, and environmental risk factors are connected with fatigue-related traffic accidents, and fatigue driving is a major cause of traffic accidents [3]. Drunk driving is one of the major behavioral issues connected with problematic alcohol consumption [4]. Speed is one of the most important factors in traffic safety as higher speeds are linked to increased crash risk and higher injury severities [5]. The young adults are easily at the highest risk of engaging in distracted driving [6]. In this paper, the state of driving risk is the behaviors during deceleration zone, which can easily cause adverse consequences.

The process of driving behavior will be presented in specific vehicle movement. The parameters of vehicle movement are generally vehicle speed, acceleration, and vehicle trajectory [7]. So, the driving risk can be identified by observing parameters of vehicle movement [8]. Many measures have been proposed to monitor the driving risk, but they mainly focus on the determination of the driver's physiological indicators [9–11]. Neural network is an important method to detect driving risk [12]. With the fast development and adoption of vehicle-to-infrastructure (V2I) facilities, it is very convenient to detect driving risk, and the accuracy of the detection model is a key factor [13–15]. So, a neural network model of high accuracy will be established to identify driving risk in this paper.

Speed is one of the most important factors in traffic safety, and on-ramps and off-ramps in expressway are typical sections [16]. The work zone in expressways has higher rear-end crash risk than that in arterials [17]. There is strong

TABLE 1: Set of characterization indicators for driving risk.

Variable name	Unit	Indicators	Description
t_R	s	Reaction time	It characterizes the zone time between the drivers' reading the speed limit sign and vehicles slowing down.
v_I	m/s	Initial speed	It characterizes the speed at which the driver finds the speed limit sign.
v_F	m/s	Final speed	It characterizes the speed which the driver can keep after the deceleration is completed.
Δv	m/s	Velocity difference	It characterizes the change in speed for the driver after seeing the speed limit sign.
a	m/s ²	Acceleration	It characterizes the rate of the speed change during deceleration.

relationship between speed compliance and the frequency and severity of traffic accidents [18]. Vehicle speed changing and driver reaction time can be measured accurately in deceleration zone of expressway [19, 20]. So, five typical parameters are selected to represent the vehicle speed changing and driver reaction time. Besides, a generalized regression neural network (GRNN) is used as a correlation method of the five typical parameters [21].

The study of driving risk detection is carried out based on GRNN model in deceleration zone of expressway, which aims to reduce the incidence of accidents and provide security for drivers' safety.

2. Influencing Factors Analysis for Driving Risk

According to the relevant research results, there are three main factors to affect the driver's driving ability [22], which are as follows:

- (1) Fatigue causes the driver's body function to be unbalanced
- (2) Distracted driving makes it difficult for drivers to concentrate their attention, due to interference from external factors
- (3) There is temporary injury of the driver, such as a decrease in the actual working ability of the driver, which is due to side effects of drugs and alcohol

The state of vehicle in the deceleration zone is a specific expression of the situation in which the driver manipulates the vehicle, and it can determine whether the driver is in a dangerous condition. An important feature of drivers in a dangerous situation is an increase in reaction time of the operating vehicle, which is manifested in the running state of the vehicle. So the reaction time of deceleration is longer when traveling through the deceleration zone. The related studies have shown that drivers who are in dangerous driving conditions often respond to sudden speed brakes to reduce the speed of the vehicle. Therefore, excessive acceleration is

a poor condition for driving. In addition, when the vehicle is traveling through the deceleration zone, if the speed is too large and the speed is not lowered than the prescribed speed, it is also a dangerous driving state [23]. So, it is useful to detect drivers' state via parameters of vehicle movement.

In order to establish the detection model for driving risk, the five indicators are initially selected as characterization indicators of the danger driving assessment, which are reaction time, acceleration, initial speed, final speed, and velocity difference. The details are given in Table 1. According to the characteristics of the deceleration zone, we designed the detection process of dangerous driving state, which is as showed in Figure 1.

3. Experimental Testing and Data Optimization

The goal of this experiment is to collect data of drivers and vehicles in the deceleration zone, such as acceleration, reaction time, initial speed, final speed, and velocity difference. It is the data foundation for constructing the detection model for driving risk.

3.1. Experimental Implementation Process. Based on the experimental purpose, we have developed a comprehensive data acquisition program via simulated driving experiments. We chose the scene of expressway deceleration zone in the simulated driving platform as the typical section, and the simulated driving platform contains multiple sensors for collecting behavior and vehicle signals, which has been applied to various driving behavior experiments [24].

There are four driving states of "normal driving," "speeding driving," "distracted driving," and "fatigue driving." Each driver continues to be tested for about 30 minutes under the four driving conditions. Experimental equipment collects drivers' physiological information, vehicle operating parameters, and driving environment information in real time. Drivers control the simulated driving platform to complete

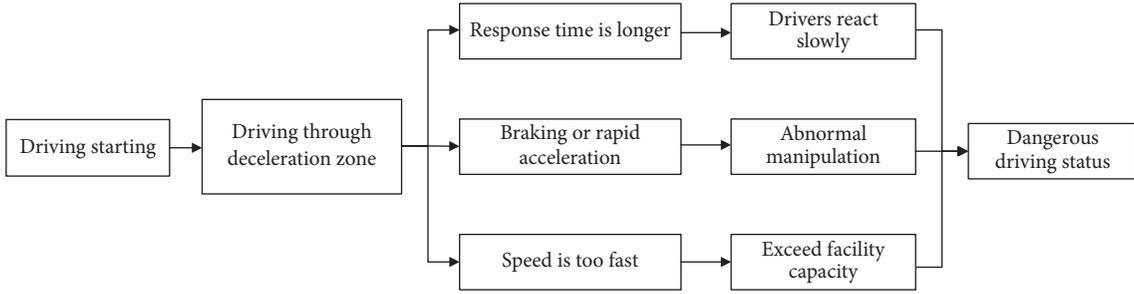


FIGURE 1: Detection process of dangerous driving state.

TABLE 2: Parts of the original sample data.

Status	Acceleration	Reaction time	Initial speed	Final speed	Velocity difference
Normal	-6.38	0.48	106	50	56
Normal	-5.66	0.24	104	51	53
Normal	-6.14	0.32	98	70	28
Danger	-2.41	0.88	115	80	35
Danger	-2.16	0.84	115	95	20
Danger	-2.28	1.08	130	110	20
...

the natural operation according to the requirements of the experimental operation process, and the multiple sensors collect five required indicators, namely, the initial speed v_I , the final speed v_F , the velocity difference Δv , the reaction time t_R , and the acceleration a .

When the experiment is completed, the data of each driver's personality, vehicle running status, and driving environment are summarized, and the driving experimental database in the expressway deceleration zone is established to complete the detection model.

3.2. Extraction for Effective Data. According to the experimental scheme, we need to obtain the information of five raw parameters via the multiple sensors, and the parameters are the initial speed v_I , the corresponding time t_1 when the drivers see the speed limit sign, the time t_2 when driver's action starts to decelerate via seeing the speed limit sign, the time t_3 when the drivers decelerate to a certain speed and are about to move at a constant speed, and the final speed v_F at time t_3 . After obtaining the above sample data, the reaction time t_R , the velocity difference Δv , and the acceleration a can be calculated according to formulas (1), (2), and (3). After the above processing, we can get the five required indicators, namely, the initial speed v_I , the final speed v_F , the velocity difference Δv , the reaction time t_R , and the acceleration a .

$$t_R = t_2 - t_1 \quad (1)$$

$$\Delta v = v_F - v_I \quad (2)$$

$$a = \frac{v_F - v_I}{t_3 - t_2} \quad (3)$$

In this simulation experiment, each driver has to perform three normal driving processes and three dangerous driving

processes. The dangerous driving process is divided into speeding driving, distracted driving, and fatigue driving. For normal driving, we stipulate that the driving speed cannot exceed the speed limit value of the relevant section in the expressway. Since the maximum speed of the expressway in China is 120 km/h, the parts of data whose speed is larger than 120 km/h have been excluded. For dangerous driving, due to the simulation characteristics and the errors in the experimental operation, we also have eliminated the invalid data. The style and composition of data are shown in Table 2.

In this experiment, we have obtained more than 60 hours of data from 31 drivers, and the uniform sampling method is used to extract 332 sets of valid data in the normal and dangerous state [25]. Specifically, there are 266 samples in the training set, 133 samples of which are in the normal state and dangerous state, respectively; there are 66 samples in the testing set, 33 samples of which are in the normal state and dangerous state, respectively.

3.3. Screening for Optimal Characterization. Table 2 is the raw data set based on multiple sensors, and each set of data contains five indicators of acceleration, reaction time, initial speed, final speed, and velocity difference. In the next step, the one-way ANOVA is used to test whether the five indicators have significant differences between normal and dangerous states, and the tested indicators are selected as the best indicators of driving risk [26]. The test results are given in Table 3.

The analysis results in Table 3 show that there are significant differences ($P < 0.05$) in the initial speed, final speed, reaction time, and acceleration between normal and dangerous states. This is to say, the above four indicators can be used as the ideal indicators of driving risk. However, the

TABLE 3: Variance analysis results of various indicators.

Index	df	MS	F	P value	F-crit
Initial speed	1	11462.93	107.65	5.54E-22	3.87
Final speed	1	14432.71	102.30	4.28E-21	3.87
Velocity difference	1	170.85	1.10	0.2944	3.87
Reaction time	1	42.68	6.87	0.0092	3.87
Acceleration	1	3.20	61.77	5.68E-14	3.87

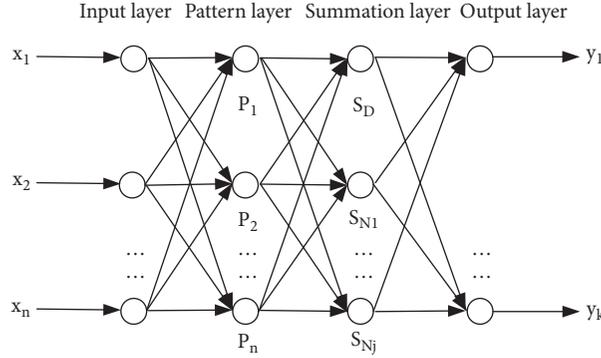


FIGURE 2: Classic structure for generalized regression neural network.

velocity difference did not pass the ANOVA testing. So, the four indicators of acceleration, reaction time, initial speed, and final speed will be used as indicators for the detection mode recognition of driving risk.

4. Detection Model of Driving Risk Based on GRNN

4.1. Structure for Classic GRNN. In many nonlinear models that solve the problem of supervised pattern recognition, the artificial neural network models are widely used due to their excellent adaptive ability. Besides, in actual modeling and computation, the GRNN exhibits higher prediction accuracy and faster computational speed than other nonlinear models. So, the GRNN model has been selected to solve this pattern recognition problem. GRNN is a feed-forward neural network model based on nonlinear regression theory, which is composed of four layers in the structure, which are input layer, pattern layer, summation layer, and output layer [27]. The classic structure of GRNN is shown in Figure 2.

The input matrix is $X = [x_1, x_2, \dots, x_n]^T$, and the output matrix is $Y = [y_1, y_2, \dots, y_k]^T$. The number of input layer neurons is equal to the dimension of the input matrix in the training sample. The input layer only sends the sample variables to the pattern layer and does not participate in the real operation. The number of pattern layer neurons is equal to the number of training set samples, and the layer neuron transfer function is

$$P_i = e^{[-(X-X_i)^T(X-X_i)/2\sigma^2]} \quad i = 1, 2, \dots, n, \quad (4)$$

The summation layer uses two types of neurons for summation. The first type computes the sum of the output lines of all pattern layer neurons. The pattern layer and each

neuron have a connection weight of 1, and the first type summation function is

$$S_D = \sum_{i=1}^n P_i, \quad (5)$$

The second type weights the neurons of all pattern layers, and the connection weight between the pattern layer and each neuron is taken as the output matrix of training set, and the second type summation function is

$$S_{Nj} = \sum_{i=1}^n (y_{ij} P_i), \quad j = 1, 2, \dots, k, \quad (6)$$

Each neuron in the output layer is associated with the two type results of the summation layer. The output function is

$$y_j = \frac{S_{Nj}}{S_D} \quad j = 1, 2, \dots, k, \quad (7)$$

The GRNN can achieve good nonlinear mapping ability and learning speed. Its weight and threshold are given in the training sample in one step. It does not have to be iterated, and the calculation amount is small.

4.2. Algorithm Process of GRNN. The GRNN has been used to construct the detection model of driving risk in deceleration zone of expressway, and the algorithm process is as follows [28].

4.2.1. Generating Training Sets and Testing Sets. To avoid generality, we use a random method to generate training sets and testing sets. As mentioned above, the experimental data includes the categories of driving status (normal driving

TABLE 4: Sample data for training sets and testing sets.

Status	Acceleration	Reaction time	Initial speed	Final speed
Normal	-6.38	0.48	106	50
Normal	-5.66	0.24	104	51
Normal	-6.14	0.32	98	70
Danger	-2.41	0.88	115	80
Danger	-2.16	0.84	115	95
Danger	-2.28	1.08	130	110
...

TABLE 5: Input variable combinations corresponding to the 15 models.

Model number	Input variable			
	Acceleration	Reaction time	Initial speed	Final speed
1	○			
2	○	○		
3	○	○	○	
4	○	○	○	○
5		○		
6		○	○	
7		○	○	○
8			○	
9			○	○
10				○
11	○		○	
12	○			○
13	○	○		○
14	○		○	○
15		○		○

and dangerous driving) and four indicators (acceleration, reaction time, initial speed, and final speed). Some sample data are shown in Table 4. Specifically, there are 266 samples in the training set and 66 samples in the testing set.

4.2.2. *Creating GRNN Model Group for Driving Risk.* We use the toolbox function “newgrnn” of MATLAB neural network to create the GRNN models, and a total combination of the four variables has been established. It can be divided into 15 groups, and each model is constructed separately. For example, the variable of input layer in model 1 is the initial speed, and the variables of input layer in model 2 are the initial speed and final speed. The combinations for input variables are shown in Table 5 corresponding to the 15 GRNN models.

4.2.3. *Simulation Testing.* After the GRNN model for driving risk is established, the input matrix of the testing set can be taken into the model, and the output of the model is the corresponding prediction result.

4.2.4. *Performance Evaluation.* By calculating the error between the prediction category and the actual category

of the testing set, the generalization ability of the GRNN model on driving risk can be evaluated. At the same time, the MATLAB function “cputime” can be used to calculate the running time of the program, which is to measure the running speed and performance of the program.

4.3. *Discussion of Results.* Since the training set and the testing set are randomly generated, the results of every calculation are slightly different. The consequences of the certain program operation have been randomly selected. So, the prediction accuracy, running time, and optimal model testing results of the model group are, respectively, shown in Figures 3, 4, and 5. After analysis, the GRNN model group of the driving risk detection in deceleration zone can clearly give the following results:

(1) It can be seen from Figure 3 that the GRNN models have good generalization ability, and the best prediction accuracy rate of testing set is up to 86.4%(2) As can be seen from Figure 3, the numbers of GRNN model established by inputting each variable alone are model 1, model 5, model 8, and model 10 (acceleration, reaction time, initial speed, and final speed). Among them, the GRNN model established by acceleration alone has the worst performance, and the



FIGURE 3: Prediction accuracy of the fifteen models.

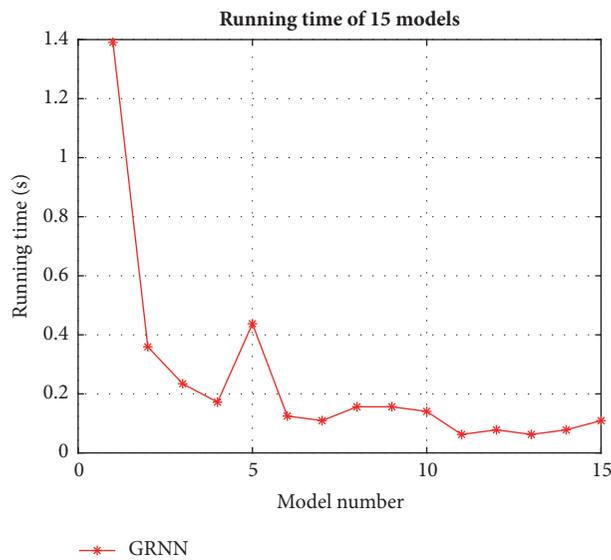


FIGURE 4: Running time of the fifteen models.

correct rate is just 59%. The GRNN models with separate input of reaction time, initial speed, and final speed have better performance, and the correct rate is, respectively, 74%, 80%, and 79%. This illustrates that the correlation between acceleration and vehicle motion is small, but the correlation between initial speed, final speed, reaction time, and vehicle motion is large

(3) In Figure 3, the input variable of model 1 is acceleration, the input variables of model 2 are acceleration and reaction time, the input variables of model 3 are acceleration, reaction time, and initial speed, and the input variables of model 4 are acceleration, reaction time, initial speed, and final speed. It can be seen from Figure 3 that the prediction accuracy is gradually improved from model 1 to

model 4, and the same rules are applied to models 5–7 and models 8–9, which indicates that multiple factors contribute to the recognition of the driving state. The number of input variables is positively correlated with the prediction accuracy rate

(4) It can be seen from Figure 4 that the average running time of the 15 models is about 0.2 s, which is much faster than the BP neural network algorithm

(5) The prediction result of testing set drawn in Figure 5 is the corresponding optimal driving risk detection for model 4, and it is a typical model in which four parameters (acceleration, reaction time, initial speed, and final speed) are involved in the operation. The prediction accuracy of model 4 is 86.4%, which is higher than the other models

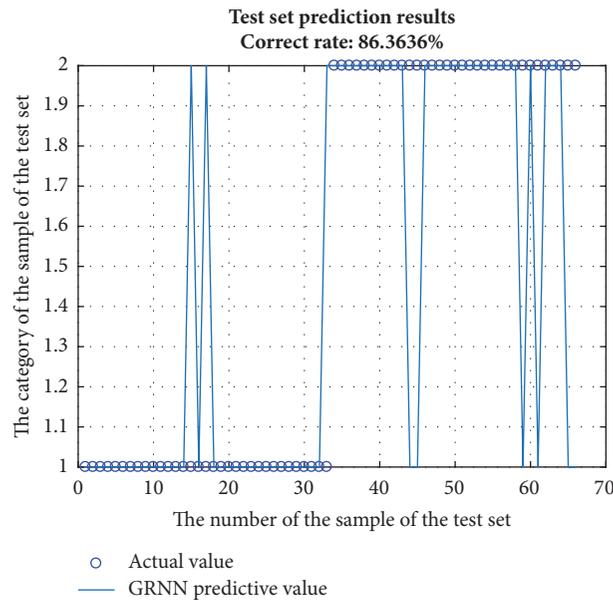


FIGURE 5: Prediction results of model 4.

5. Conclusion

One of the negatively influential issues regarding social life is a traffic accident. It has negative impacts on people's lives and well-being due to increasing their travel pressure. It has to be mentioned that it would not only threaten individuals' health, but also disturb the management system because of economic loss.

The study aims to develop driving risk detection model of deceleration zone in expressway. In this research, it is considered to increase prediction accuracy for the GRNN model. The four parameters have been selected via the ANOVA analysis. So, it is clear that number of input variables is positively correlated with the prediction accuracy rate. For this purpose, fifteen models are proposed to choose the best prediction model. Therefore, the risk of driving would be decreased remarkably if the GRNN model is used to detect vehicle movement.

In the future work, there are two aspects to improve. Firstly, more parameters should be collected and input to the GRNN model, in order to heighten the prediction accuracy rate. Then, data from actual traffic conditions are available to validate the simulated data.

Data Availability

The "variance analysis results of various indicators" data used to support the findings of this study are included within the article. The "sample data for training sets and testing sets" data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 12 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study is supported by the National Natural Science Foundation of China (NO. 71701070), the Natural Science Foundation of Guangdong Province (NO. 2016A030310427), the Science and Technology Project of Guangzhou City (NO. 201804010466), and the China Scholarship Council.

References

- [1] X. Qu, Q. Meng, and Z. Liu, "Estimation of number of fatalities caused by toxic gases due to fire in road tunnels," *Accident Analysis & Prevention*, vol. 50, pp. 616–621, 2013.
- [2] C. Wang and N. Stamatiadis, "Sensitivity analysis on new simulation-based conflict metrics," *Safety Science*, vol. 82, pp. 399–409, 2016.
- [3] G. Zhang, K. K. W. Yau, X. Zhang, and Y. Li, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Analysis & Prevention*, vol. 87, pp. 34–42, 2016.
- [4] I. Taymur, E. Budak, V. Duyan, B. B. Kanat, and S. Önen, "Examination of personality traits and social problem-solving skills of individuals whose driving licenses have been confiscated due to drunk driving," *Traffic Injury Prevention*, vol. 18, no. 1, pp. 3–8, 2017.
- [5] C. D. Fitzpatrick, S. Rakasi, and M. A. Knodler, "An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression," *Accident Analysis & Prevention*, vol. 98, pp. 57–63, 2017.
- [6] K. A. Braitman and A. L. Braitman, "Patterns of distracted driving behaviors among young adult drivers: Exploring relationships with personality variables," *Transportation Research*

- Part F: Traffic Psychology and Behaviour*, vol. 46, pp. 169–176, 2017.
- [7] Z. Liu, W. Yi, S. Wang, and J. Chen, “On the Uniqueness of User Equilibrium Flow with Speed Limit,” *Networks and Spatial Economics*, vol. 17, no. 3, pp. 763–775, 2017.
- [8] C. Wang, C. Xu, J. Xia, Z. Qian, and L. Lu, “A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation,” *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 281–291, 2018.
- [9] R. Sayed and A. Eskandarian, “Unobtrusive drowsiness detection by neural network learning of driver steering,” *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 215, no. 9, pp. 969–975, 2001.
- [10] M. V. M. Yeo, X. P. Li, K. Q. Shen, and E. P. V. Wilder-Smith, “Can SVM be used for automatic EEG detection of drowsiness during car driving?” *Safety Science*, vol. 47, no. 1, pp. 115–124, 2009.
- [11] M.-H. Sigari, M. Fathy, and M. Soryani, “A driver face monitoring system for fatigue and distraction detection,” *International Journal of Vehicular Technology*, vol. 2013, 2013.
- [12] M. Patel, S. K. L. Lal, D. Kavanagh, and P. Rossiter, “Applying neural network analysis on heart rate variability data to assess driver fatigue,” *Expert Systems with Applications*, vol. 38, no. 6, pp. 7235–7242, 2011.
- [13] M. Zhou, X. Qu, and X. Li, “A recurrent neural network based microscopic car following model to predict traffic oscillation,” *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 245–264, 2017.
- [14] M. Zhou, X. Qu, and S. Jin, “On the Impact of Cooperative Autonomous Vehicles in Improving Freeway Merging: A Modified Intelligent Driver Model-Based Approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1422–1428, 2017.
- [15] Z. Liu, Y. Zhang, S. Wang, and Z. Li, “A Trial-and-Error Method with Autonomous Vehicle-to-Infrastructure Traffic Counts for Cordon-Based Congestion Pricing,” *Journal of Advanced Transportation*, vol. 2017, 2017.
- [16] X. Qu, Y. Yang, Z. Liu, S. Jin, and J. Weng, “Potential crash risks of expressway on-ramps and off-ramps: a case study in Beijing, China,” *Safety Science*, vol. 70, pp. 58–62, 2014.
- [17] Q. Meng and J. Weng, “Evaluation of rear-end crash risk at work zone using work zone traffic data,” *Accident, Analysis & Prevention*, vol. 43, no. 4, pp. 1291–1300, 2011.
- [18] M. Matowicki and O. Pribyl, “Speed compliance in freeway variable speed limit system -case study of the Prague city ring,” *Transport Problems*, vol. 11, no. 1, pp. 69–78, 2016.
- [19] X. Qu, J. Zhang, and S. Wang, “On the stochastic fundamental diagram for freeway traffic: Model development, analytical properties, validation, and extensive applications,” *Transportation Research Part B: Methodological*, vol. 104, pp. 256–271, 2017.
- [20] X. Qu, S. Wang, and J. Zhang, “On the fundamental diagram for freeway traffic: a novel calibration approach for single-regime models,” *Transportation Research Part B: Methodological*, vol. 73, pp. 91–102, 2015.
- [21] Z. Xu, T. Wei, S. Easa, X. Zhao, and X. Qu, “Modeling Relationship between Truck Fuel Consumption and Driving Behavior Using Data from Internet of Vehicles,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 3, pp. 209–219, 2018.
- [22] P. A. Hoggarth, C. R. H. Innes, J. C. Dalrymple-Alford, and R. D. Jones, “Prediction of driving ability: Are we building valid models?” *Accident Analysis & Prevention*, vol. 77, pp. 29–34, 2015.
- [23] K. A. Suhr and C. S. Dula, “The dangers of rumination on the road: Predictors of risky driving,” *Accident Analysis & Prevention*, vol. 99, pp. 153–160, 2017.
- [24] W. Qi, H. Wen, Y. Wu, and L. Qin, “Effect model of urban traffic congestion on driver’s lane-changing behavior,” *Advances in Mechanical Engineering*, vol. 9, no. 9, pp. 1–12, 2017.
- [25] A. P. Larocca, R. L. Ribeiro, A. da Cruz Figueira, P. T. de Oliveira, L. C. Lulio, and M. A. Rangel, “Analysis of perception of vertical signaling of highways by drivers in a simulated driving environment,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 58, pp. 471–487, 2018.
- [26] J. Dien, “Best practices for repeated measures ANOVAs of ERP data: Reference, regional channels, and robust ANOVAs,” *International Journal of Psychophysiology*, vol. 111, pp. 42–56, 2017.
- [27] B. M. Ozyildirim and M. Avci, “Generalized classifier neural network,” *Neural Networks*, vol. 39, pp. 18–26, 2013.
- [28] Ö. Polat and T. Yıldırım, “Genetic optimization of GRNN for pattern recognition without feature extraction,” *Expert Systems with Applications*, vol. 34, no. 4, pp. 2444–2448, 2008.

Research Article

Modelling Rail-Based Park and Ride with Environmental Constraints in a Multimodal Transport Network

Xinyuan Chen and Inhi Kim 

Monash Institute of Transport Studies, Australia

Correspondence should be addressed to Inhi Kim; inhi.kim@monash.edu

Received 3 July 2018; Revised 4 September 2018; Accepted 12 September 2018; Published 4 October 2018

Guest Editor: David Z. W. Wang

Copyright © 2018 Xinyuan Chen and Inhi Kim. 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.

Environmental sustainability is a significant aspect in the sustainable development of modern urban cities, especially in the road transport system. As traffic demands increase, public transport requires more promotion to accommodate the increasing travel demands while maintaining the environmental quality. Public transport, however, is less attractive in vast suburb areas mainly due to its longer travel distance and waiting time. Therefore, this paper proposes a rail-based Park-and-Ride (RPR) scheme to promote public transport in the multimodal transport network. To remedy the heterogeneous distribution of vehicle pollutants in the network, regulations in environmental sensitive districts are required and studied in this paper. To quantitatively evaluate and analyse this joint RPR and environmental regulation strategy in multimodal transport systems, this paper develops an environmental constrained combined modal split and traffic assignment (EC-CMSTA) model. The proposed formulation adopts the concept of fix-point to reformulate the nonlinear complementarity conditions associated with the combined modal split and user equilibrium conditions, which is subsequently incorporated into a VI formulated nonlinear complementarity conditions associated with environmental constraints. The proposed VI formulation can handle a general constraint structure, which enhances the modelling adaptability and flexibility. The strictly monotone and Lipschitz continuity properties of this model are rigorously proved, giving rise to efficient algorithms for the model. A customized projection based self-adaptive gradient projection (SAGP) algorithm is then developed. Numerical studies demonstrate that the EC-MSTA model could enhance the behavioural modelling of network users' travel decisions and assist in quantitatively evaluating the effectiveness of RPR schemes and environmental regulations.

1. Introduction

Environmental protection is an important component in the sustainable development of modern urban cities and closely related to the transport system. According to Hong Kong Air Pollutant Emission Inventory (2016), the amount of CO generated in the year of 2016 in Hong Kong is 58600 tones and among them, road transport contributes to 31500 tones and shares 54% of the CO emissions. Also, the US's environmental protection agency reported motorized vehicles in the United States contribute to 86% of CO emissions, 62% of NO_x emissions, 12% of PM_{2.5} emissions, and 7% of SO₂ emissions [1]. Kitthamkesorn et al. [2] pointed out that only about 20% of the world's town residents enjoy good enough air quality as measured by the levels of emissions and over 600 million people in urban areas worldwide were exposed to dangerous

levels of traffic-generated emissions from congestion. In addition, these emissions have not only adverse effects on human health but also a source of greenhouse gas that contributes to global warming [3].

With these adverse effects, the promotion of alternative green transport modes is thus pressing for the governments and transportation authorities to accommodate the growing travel demand while maintaining the environmental quality. Public transport aiming to satisfy the travel demand of urban commuters safely, quickly, and affordably is therefore a significant alternative to private transport. However, in vast suburb areas which are far away from an urban centre, public transport becomes less attractive, mainly due to the unreliable time table, tortuous bus itinerary, much lower travel speed, and lack of door to door services [4]. Large-scale deployments of public transport system in the remote areas are hardly seen

as an economical available strategy. Park and Ride (P&R), therefore, becomes an important strategy to promote public transport since it enables travellers to drive when the road is not congested and to change a mode to transit accessing to CBD without traffic congestion. Since its first implementation in 1930s, P&R has been widely recognized as a significant traffic management strategy and becomes an indispensable part of majority modern urban cities.

P&R can be categorized into bus-based P&R (BPR) and rail-based P&R (RPR) based on its ride component [5]. In a metropolis with advanced rail system, rail-based P&R is superior to bus-based P&R, due to the faster speed, higher level of safety, comfortability, and its advantages in environmental protection (zero-emission).

Despite the fact that environmental deterioration could somehow be alleviated by promoting public transport through RPR schemes, the environmental requirements are usually not strictly met in certain districts mainly due to the heterogeneous distribution of pollution in the network [3]. To remedy this limitation, specific environmental regulations, therefore, are required to certain district, e.g., the city business district (CBD) with many tall buildings, to protect the environmental quality in this area.

Therefore, we aim to use RPR services to assist mode shift to reduce the overall vehicle travel demands in the road network and then apply area-based regulation scheme to deal with the environmental deterioration caused by heterogeneity of traffic flows. The environmental protection approaches are integrated as a combined strategy in this paper. Previous studies revealed that a systematic study is necessary to properly understand and evaluate the effectiveness of the environmental regulations, as well as the RPR scheme. Thus, this paper focuses on the modelling and solution algorithm development of the proposed environmental protection strategy.

2. Literature Review

Road network equilibrium approaches are commonly preferred for estimating vehicle emissions and noises [3], because the distribution of vehicle pollution heavily relies on the states of traffic flows [6]. Other than the overall pollution level, the spatial distribution level should be controlled since it is heterogeneous.

One line of the research developed models by setting environmental protection as objectives [3], modelling environmental considerations as an objective function. Rilett and Benedek [7] proposed the concept of emission optimum (EO) and emission-oriented user equilibrium (EOUE), in which the former one assumes travellers make route choices with the objective of minimizing the total emissions or noises, and the latter includes the toll charges to represent the impact on the environment. Later, Benedek and Rilett [8] made extension by modelling the assignment of vehicles based on equity which considers the objectives of neighbouring residents, rather than the system operator's or drivers' objectives. Sugawara and Niemeier [9] used the emissions-optimized trip assignment model to estimate the maximum carbon monoxide emission reduction under varying congestion levels.

Although the overall environmental quality could somehow be improved by taking the environmental considerations

as the objective, the pollution level is not always strictly met on each link and zone in the road network. Therefore, another line of the research focuses on modelling the emission as environmental constraints. Chen, Zhou and Ryu [10] developed a traffic assignment model by considering various physical and environment restrictions as side constraints, which is formulated as a VI problem and solved by a predictor-corrector (PC) algorithm. Li et al. [11] developed a toll design model by simultaneously considering congestion and environmental externalities. Recently, Xu, Chen, and Cheng [1] reformulated the environmental constrained traffic assignment model through a smooth gap function to enhance the modelling adaptability and flexibility.

As a rational extension of traffic assignment problem, Nagurney, Ramanujam, and Dhanda [12] extended the environmental constrained traffic assignment problem to the case of multimodal traffic network. However, the model presented in this paper simply follow the deterministic user equilibrium (DUE) principle for both modal split and traffic assignment. In order to evaluate the environmental regulation strategy in the multimodal transport network, a better behavioural model is needed. Therefore, the combined modal split and traffic assignment model (CMSTA) is adopted in this paper. In the literature, the CMSTA problems are widely adopted to evaluate the P&R schemes. Fernandez et al. [13] conducted an initial work to model P&R in deterministic case. Later, this work was extended to the case of asymmetric link travel time functions by García and Marín [14] and to the logit-based stochastic model by Li et al. [15] and Lam et al. [16]. Recently, Liu and Meng [5] provided a formulation for the bus-based P&R in a network with the congestion charge. Pineda et al. [17] and Song, He, and Zhang [18] developed an integrated P&R facilities and transit service optimization model. Kitthamkesorn et al. [2] and Kitthamkesorn and Chen [19] considered the influence of "go green" transport modes on environmental protection by taking use of recent advances in logit model. Liu et al. [20] proposed a new concept of P&R, remote park and ride, and developed an exact solution method for the optimal design of parking facilities, following the nonlinear valid inequality technique [21, 22]. Some other studies focus on the P&R schemes in a linear corridor, e.g., Liu et al. [23], Wang and Du [24], and Du and Wang [25]. However, these studies mainly focus on congestion alleviation. As abovementioned, the environmental requirements in certain areas are not considered. Therefore, a more general modelling framework that incorporates environmental requirements is needed. To this end, this paper aims to build an environmental constrained combined modal split and traffic assignment (EC-CMSTA) model to better evaluate and analyse environmental sustainable transportation development strategies.

This paper distinguishes itself from previous studies in threefold aspects. First, we propose EC-CMSTA, CMSTA model with side constraints, which could handle the general environmental constraint structures, including linear or nonlinear, link based, or area-based constraints. Also, the general link/route cost structure is considered here. Second, by incorporating general side constraints into CMSTA, solving is much more complex compared to the traditional side constrained unimodal traffic equilibrium problems. To explicitly

consider the general environmental constraints and make use of existing efficient algorithms, we propose a VI formulation by adapting the concept of fix-point. This VI formulation is strictly monotone and Lipschitz continuous, permitting a number of existing efficient algorithms for its solution [10, 26]. Third, even though the solution methods for the traffic assignment problems with side constraints have been well studied with numerous achievements [1, 10, 12, 27], solutions for the multimodal traffic equilibrium with side constraints are relatively limited. Recently, Meng, Liu and Wang [28] proposed a CA incorporated self-adaptive PC algorithms to solve asymmetric stochastic user equilibrium (SUE) problem with elastic demand and link capacity constraints. Inspired by Meng, Liu, and Wang [28], we extend this approach to multimodal traffic equilibrium problem and develop a customized projection based algorithm with a self-adaptive step size scheme which permits the convergence under a much milder assumption of the underlying mapping.

The remainder of this paper is organized as follows: we present assumptions and an EC-CMSTA model for the environmental regulations in the multimodal transport system with RPR scheme in the next section. A general VI formulation and a projection based solution algorithm are provided in Sections 3 and 4. Section 5 provides two numerical studies to validate the proposed methodology. Finally, Section 6 concludes this study.

3. Problem Description

Considering a specific environmental regulation strategy, to analyse its impact in the multimodal transport system, we build a model for the equilibrium flows following environmental requirements.

Given a multimodal transportation network represented by a strongly connected topology graph $G(N, A)$ where N is the set of nodes and A is the set of links. The origins, destinations, and RPR facilities are scattered in the network and represented by certain nodes, $n \in N$. The set of links A can be further categorized into three subsets: set of autolinks A^a ; set of rail links A^r ; and set of P&R links A^p ; $A = A^a \cup A^r \cup A^p$.

3.1. Modelling Environmental Requirements. As abovementioned, the environmental requirements are not strictly met in the road network. Thus, it is necessary to explicit considering environmental protection thresholds to prevent environmental deterioration in certain zones. In the literature, the environmental constraints can be expressed into a general form:

$$g_e(\mathbf{v}) \leq 0, \quad \forall e \in E \quad (1)$$

where e is a component of E and E represents a set of environmental sensitivity elements. In the case of link based environmental protection strategy, e represents a link which is assumed in this paper. \mathbf{v} is the vector of link flows; $g_e(\mathbf{v})$ is a function to measure the amount of vehicle pollution and environmental thresholds. Equations (1) ensure the amount of pollution no larger than predetermined thresholds. Note that (1) are general formulation that could measure different types of vehicle pollutants [1]. Herein, we make assumption as follows.

Assumption 1. The environmental threshold measure function $g_e(\mathbf{v}), \forall e \in E$, is a continuously differentiable and strictly monotone increasing function of the link flows \mathbf{v} .

Suppose the total number of environmentally sensitive areas in the network is I . The inbound flows of these areas should be restricted according to environmental protection strategies. All the links in the area $i \in I$ are grouped into set E_i , where $E_i \in A$. Then, the specific link based environmental constraints in these areas could be expressed as

$$g_a(\mathbf{v}) \cdot v_a \leq \bar{g}_a, \quad \forall a \in E_i, i \in I \quad (2)$$

where $g_a(\mathbf{v})$ is the amount of vehicle pollutants per vehicle on link a and \bar{g}_a is the environmental protection threshold on link a imposed by the operators.

3.2. Network Assessment. Due to the fact that the travel decisions of a network user consist of modal choice and route choice, a combined modal split and traffic assignment model is needed. The travel impedance on an auto or rail link is clearly influenced by its flows. For the RPR links, each of them always connects an autolink and a rail link, so that the RPR flows could transfer. The total travel impedance on a RPR link includes the time spent on parking, waiting/boarding the dedicated bus as well as the bus in-vehicle travel time, which is clearly flow-dependent. We then make the following assumption.

Assumption 2. The link travel time $t_a(\mathbf{v}), a \in A$ is a continuously differentiable, monotone increasing of the link flow \mathbf{v} .

Let K_m^{od} denote the set of paths under m ; $f_k^{m,od}$ denotes the flow on path $k \in K_m^{od}$ for mode m ; q_m^{od} denotes the demand of mode m ; and \bar{q}^{od} denotes the total travel demand between OD pair (o, d) . We have the following flow conservation conditions:

$$\sum_{m \in M_{od}} q_m^{od} = \bar{q}^{od}, \quad \forall od \in W \quad (3)$$

$$q_m^{od} \geq 0, \quad \forall m \in M_{od}, od \in W \quad (4)$$

$$\sum_{k \in K_m^{od}} f_k^{m,od} = q_m^{od}, \quad \forall m \in M_{od}, od \in W \quad (5)$$

$$f_k^{m,od} \geq 0, \quad \forall k \in K_m^{od}, m \in M_{od}, od \in W \quad (6)$$

$$v_a = \sum_{od \in W} \sum_{m \in M_{od}} \sum_{k \in K_m^{od}} f_k^{m,od} \delta_{a,k}^{m,od}, \quad \forall a \in A \quad (7)$$

where (3) and (5) define demand/flow conservation conditions. Equations (4) and (6) are the nonnegativities. Equations (7) define the link-path flow relationship. Constraints (1), (3)-(7) define the feasible set of Ω of $(\mathbf{v}, \mathbf{f}, \mathbf{q})$ where $\mathbf{v} := (v_a, a \in A)$ is the vector of link flows; $\mathbf{f} := (f_k^{m,od}, k \in K_m^{od}, m \in M^{od}, od \in W)$ is the vector of path flows; and $\mathbf{q} := (q_m^{od}, m \in M^{od}, od \in W)$ is the vector of modal demands.

Let W denote the set of OD pairs; \bar{q}^{od} denote the total travel demand for $(o, d) \in W$; and M_{od} denote the set of modes between OD pair (o, d) . The multinomial logit (MNL) model is adopted to analyse the modal split pattern. The probability of choosing any mode m between an OD pair (o, d) equals

$$p_m^{od} = \frac{\exp(-\theta \cdot \tau_m^{od})}{\sum_{i \in M_{od}} \exp(-\theta \cdot \tau_i^{od})}, \quad \forall m \in M, od \in W \quad (8)$$

where $\theta \in [0, +\infty)$ is the dispersion parameter and τ_m^{od} is the deterministic component of the utility of mode m between OD pair (o, d) , which can be taken as the equilibrium path travel time of mode m . The satisfaction of MNL, $S_{od}(\tau_m^{od})$, equals $\tau_m^{od} + (1/\theta) \ln(q_m^{od})$, which is concave function [28].

The user equilibrium path flow solution is then formulated by the following conditions:

$$f_k^{m,od} (c_k^{m,od} - \varphi_m^{od}) = 0, \quad \forall k \in K_m^{od}, m \in M_{od}, od \in W \quad (9)$$

$$\begin{aligned} f_k^{m,od} &\geq 0, \\ (c_k^{m,od} - \varphi_m^{od}) &\geq 0, \end{aligned} \quad \forall k \in K_m^{od}, m \in M_{od}, od \in W \quad (10)$$

where φ_m^{od} is the Lagrange multiplier associated with the constraints (5), which represents the travel impedance of using mode m . At equilibrium, the travel time $c_k^{m,od}$ of a used path $k \in K_m^{od}$ equals φ_m^{od} . All commuters under mode m have identical path travel time, and no traveller can unilaterally reduce his/her travel time by changing routes.

4. Variational Inequality Formulation

In this section, we propose two equivalent VI formulations for the proposed EC-CMSTA model. In order to strengthen the flexibility and adaptability of the proposed model, we further make the following two assumptions.

Assumption 3. The link travel time function is asymmetric, i.e., t_a is not only affected by its own flow v_a , but also flows on some other links, which is expressed as $t_a = t_a(\mathbf{v})$.

Assumption 4. The path travel time c_k is nonadditive to its links, i.e., $c_k \neq \sum_a t_a \delta_{a,k}$, where $\delta_{a,k} = 1$ if path k uses link $a \in A$ and 0 otherwise.

Then, the VI mode for EC-CMSTA is formulated as follows:

$$\begin{aligned} &\sum_{od \in W} \sum_{m \in M_{od}} \sum_{k \in K_m^{od}} c_k^{m,od}(\mathbf{f}^*, \mathbf{q}^*) (f_k^{m,od} - f_k^{m,od*}) \\ &+ \sum_{od \in W} \sum_{m \in M_{od}} \bar{c}_m^{od}(\mathbf{f}^*, \mathbf{q}^*) (q_m^{od} - q_m^{od*}) \geq 0, \end{aligned} \quad (11)$$

$$\forall (\mathbf{f}^*, \mathbf{q}^*) \in \Omega$$

where $c_k^{m,od}$ is the travel time on route k , where $k \in K_m^{od}$. $\bar{c}_m^{od} = (1/\theta) \ln(q_m^{od})$ represents the entropy term in travel cost

function of mode m , $m \in M_{od}$; $f_k^{m,od*}$ and q_m^{od*} are the optimal solution of this problem; \mathbf{f}^* and \mathbf{q}^* are the vector of optimal solution.

Let \mathbf{c} be the vector of $c_k^{m,od}$ and $\bar{\mathbf{c}}$ be the vector of \bar{c}_m^{od} , i.e., $\mathbf{c} := (c_k^{m,od}, k \in K_m^{od}, m \in M_{od}, od \in W)^T$ and $\bar{\mathbf{c}} := (\bar{c}_m^{od}, m \in M_{od}, od \in W)^T$. The VI formulation can be simplified as

$$\begin{aligned} \mathbf{c}(\mathbf{f}^*, \mathbf{q}^*)^T (\mathbf{f} - \mathbf{f}^*) + \bar{\mathbf{c}}(\mathbf{f}^*, \mathbf{q}^*)^T (\mathbf{q} - \mathbf{q}^*) &\geq 0, \\ \forall (\mathbf{f}^*, \mathbf{q}^*) &\in \Omega \end{aligned} \quad (12)$$

Note that the inclusion of nonlinear environmental constraints makes the feasible set more complex in general. Existing solution algorithms for CMSTA cannot be directly adopted to solve it. Most of existing studies focus on solving the simplified linear link capacity constrained traffic equilibrium problems by a penalty based algorithm. However, due to the complexity of EC-CMSTA (simultaneous achieving equilibrium for both route flow and modal demand), reformulating it into an integrated formulation or directly solving it by a penalty/gradient based algorithm is less effective. Therefore, this paper develops an alternative approach to reformulate the EC-CMSTA which permits existing solution algorithms for its solution.

4.1. Alternative VI Formulation. Daganzo [29] built a fix-point model in terms of link flows for the asymmetric SUE problem with fixed demand and showed availability of the MSA, which was extended by Cantarella [30] to the case of asymmetric SUE problem with elastic demand. Recently, Meng, Liu and Wang [28] further made extension by incorporating link capacity constraints. This problem was formulated as a VI problem by incorporating the fix-point formulated asymmetric SUE problem with elastic demand. Following the same technique, we propose a fixed point reformulated CMSTA model which is expressed as follows:

$$\begin{aligned} v_a &= \sum_{od \in W} \bar{q}^{od} \sum_{m \in M_{od}} p_m^{od}(\mathbf{f}(\mathbf{v})) \times P_a^{m,od}(\mathbf{f}(\mathbf{v})), \\ &\forall a \in A \end{aligned} \quad (13)$$

where \mathbf{v} is the equilibrium link flow solution, p_m^{od} is the probability of mode choice defined in (8), and $P_a^{m,od}$ is the link usage probability defined by

$$P_a^{m,od}(\mathbf{f}(\mathbf{v})) = \sum_{k \in K_m^{od}} p_k^{m,od}(\mathbf{f}(\mathbf{v})) \delta_{k,a}^{m,od}, \quad \forall a \in A \quad (14)$$

and $p_k^{m,od}$ is the probability of path k perceived as the shortest one among the path set K_m^{od} . In the deterministic case, $p_k^{m,od}$ equals one if path k is the shortest path among alternatives and zero otherwise.

A set of generalized SUE conditions with elastic demand and capacity constraints was proposed by Meng, Liu, and Wang [28] for the optimal solution of SUE problem with elastic demand, asymmetric link travel time function and link capacity constraints. These conditions can be further

extended to the multimodal transport system which is the problem addressed in this paper. Let \mathbf{v} denote the equilibrium link flow pattern of the addressed problem if and only if there exists a vector of Lagrangian multipliers corresponding to environmental constraints, denoted by $\mathbf{u}^* = (u_e^*, e \in E)^T$.

$$v_a(\mathbf{u}) = \sum_{od \in W} \sum_{m \in M_{od}} q_m^{od}(\mathbf{f}(\mathbf{v}), \boldsymbol{\lambda}(\mathbf{u}^*)) \quad (15)$$

$$\times P_a^{m,od}(\mathbf{f}(\mathbf{v}), \boldsymbol{\lambda}(\mathbf{u}^*)), \quad \forall a \in A$$

$$-g_e(\mathbf{v}) \geq 0, \quad \forall e \in E \quad (16)$$

$$u_e^* \cdot (-g_e(\mathbf{v})) = 0, \quad \forall e \in E \quad (17)$$

$$u_e^* \geq 0, \quad \forall e \in E \quad (18)$$

where $\boldsymbol{\lambda}(\mathbf{u}^*)$ is defined by

$$\boldsymbol{\lambda}(\mathbf{u}^*) = \left(\lambda_k^{m,od} = \sum_{e \in E} u_e^* \zeta_{e,k}^{m,od}, \quad \forall k \in K_m^{od}, m \in M_{od}, od \in W \right)^T \quad (19)$$

and $\zeta_{e,k}^{m,od} = 1$ if path $k \in K_m^{od}$ goes through $e \in E$ and $\zeta_{e,k}^{m,od} = 0$, otherwise. Based on the optimal Lagrangian multiplier \mathbf{u}^* , the EC-CMSTA is identical to the CMSTA in terms of generalized link travel time functions.

$$\hat{t}_a(\mathbf{v}, \mathbf{u}^*) = \begin{cases} t_a(\mathbf{v}) + u_a^*, & \forall a \in E \\ t_a(\mathbf{v}), & \forall a \in A/E \end{cases} \quad (20)$$

Equations (16)-(18) imply that if there is no environmental constraint on any link $a \in E$, $u_a^* = 0$.

Note that the complementary slackness conditions (see (16)-(18)) could be regarded as a nonlinear complementary problem (NCP) of the vector of Lagrangian multipliers \mathbf{u}^* . Due to the feasible set of \mathbf{u} is defined on the whole nonnegative orthant, this NCP problem can be equivalently formulated as a VI model. Hence, the alternative formulation of VI (11) is proposed as follows:

$$\mathbf{g}(\mathbf{u}^*)^T (\mathbf{u} - \mathbf{u}^*) \geq 0, \quad \forall \mathbf{u} \in \mathfrak{R}_+^{|E|} \quad (21)$$

where $\mathfrak{R}_+^{|E|} = (u_e \geq 0, e \in E)^T$ denotes the feasible set. $\mathbf{g}(\mathbf{u})$ is the $|E|$ dimensional vector functions, which measures vehicle pollution:

$$\mathbf{g}(\mathbf{u}) = (-g_e(\mathbf{v}(\mathbf{u})), \quad \forall e \in E)^T : \mathfrak{R}_+^{|E|} \longrightarrow \mathfrak{R}^{|E|} \quad (22)$$

$\mathbf{g}(\mathbf{u}^*)$ is thus the optimal solution of the VI problem (21). Herein, $\mathbf{v}(\mathbf{u})$ is a link flow solution for the CMSTA with the generalized link travel time functions (20) (without environmental constraints). In other words, $\mathbf{v}(\mathbf{u})$ is a solution of the fix-point model with the Lagrangian multiplier \mathbf{u} :

$$v_a(\mathbf{u}) = \sum_{od \in W} \sum_{m \in M_{od}} q_m^{od}(\hat{\mathbf{t}}(\mathbf{v}, \mathbf{u})) \times P_a^{m,od}(\hat{\mathbf{t}}(\mathbf{v}, \mathbf{u})), \quad (23)$$

$$\forall a \in A$$

where $\hat{\mathbf{t}}(\mathbf{v}, \mathbf{u}) = (\hat{t}_a(\mathbf{v}, \mathbf{u}), \quad \forall a \in A)$ denotes the vector of generalized link travel time functions.

Note that the feasible set $\mathfrak{R}_+^{|E|}$ of $VI(\mathbf{g}, \mathfrak{R}_+^{|E|})$ is unbounded, which makes it difficult to shown the existence of a solution to VI model $VI(\mathbf{g}, \mathfrak{R}_+^{|E|})$. Fortunately, according to Meng, Liu and Wang [28], one can always find an upper bound \bar{M} for $u_a, \forall a \in E$. Therefore, we define a subset $\bar{\Omega}$ of $\mathfrak{R}_+^{|E|}$ which is bounded by \bar{M} . The restricted VI model is subsequently formulated as

$$\mathbf{g}(\mathbf{u}^*)^T (\mathbf{u} - \mathbf{u}^*) \geq 0, \quad \forall \mathbf{u} \in \bar{\Omega} \quad (24)$$

The feasible set $\bar{\Omega}$ is a nonempty, closed, and convex set. To show the existence of solution to $VI(\mathbf{g}, \bar{\Omega})$ as well as the convergence of its solution algorithm, the following properties of vector function $\mathbf{g}(\mathbf{u})$ are proposed. Rigorous proofs of Propositions 5 and 6 are given in the Appendices A and B, respectively.

Proposition 5. *Vector function $\mathbf{g}(\mathbf{u})$ is strictly monotone on $\bar{\Omega}$; namely,*

$$(\mathbf{g}(\mathbf{u}') - \mathbf{g}(\mathbf{u}''))^T (\mathbf{u}' - \mathbf{u}'') > 0, \quad \forall \mathbf{u}', \mathbf{u}'' \in \bar{\Omega} \quad (25)$$

Proposition 6. *Vector function $\mathbf{g}(\mathbf{u})$ is Lipschitz continuous on $\bar{\Omega}$; namely,*

$$\|\mathbf{g}(\mathbf{u}') - \mathbf{g}(\mathbf{u}'')\|_2 \leq L \|\mathbf{u}' - \mathbf{u}''\|, \quad \forall \mathbf{u}', \mathbf{u}'' \in \bar{\Omega} \quad (26)$$

Though we have Propositions 5 and 6, the existence of a solution to VI model $VI(\mathbf{g}, \bar{\Omega})$ can be ensured. We proceed to prove the equivalence between the VI model $VI(\mathbf{g}, \bar{\Omega})$ and the generalized CMSTA conditions and summarized in Proposition 7. Rigorous proofs are given in Appendix C.

Proposition 7. *$\mathbf{u}^* = (u_a^*, \forall a \in E)$ is a solution of $VI(\mathbf{g}, \bar{\Omega})$ if and only if \mathbf{u}^* and $\mathbf{v}(\mathbf{u}^*)$ fulfil the generalized CMSTA conditions (15)-(18).*

5. Solution Algorithm

As the vector function $\mathbf{g}(\mathbf{u})$ of the VI model $VI(\mathbf{g}, \bar{\Omega})$ is an implicit function, it is burdensome to directly evaluate its gradient. On the other hand, with any given Lagrangian multiplier vector \mathbf{u} , the value of vector function $\mathbf{g}(\mathbf{u})$ can be evaluated by solving the CMSTA with generalized travel time with no constraints, which can be efficiently solved by a number of existing solution algorithms.

Projection type algorithms have been thoroughly studied and proven to be efficient in solving traffic equilibrium problems [31–33]. Among them, the gradient projection (GP) is commonly adopted due to its concise form and good performance, for which a recent survey can be found in Ryu, Chen, and Choi [34]. The self-adaptive gradient projection (SAGP) algorithm was originally proposed by He et al. [35], where the gradient projection operation is adopted to establish the descent direction, coupled with a self-adaptive step size scheme to permit the convergence under a much milder assumption of the underlying mapping, and thus can handle general cost mapping structures and generalized side

constraints. Later, Han and Sun [36] adopted this method to solve the asymmetric traffic assignment problems. Chen, Zhou, and Xu [32] further used this method to solve the asymmetric traffic assignment problem with nonadditive route cost functions. Inspired by previous studies, this paper presents the SAGP algorithm specifically designed for solving the CMSTA with generalized travel time functions.

As for the Lagrangian multipliers \mathbf{u} , the projection operation on Ω of the VI model $VI(\mathbf{g}, \bar{\Omega})$ is straightforward, as well. Propositions 5 and 6 imply the monotony and continuity of the VI model $VI(\mathbf{g}, \bar{\Omega})$. Therefore, the projection based algorithm is also suitable to evaluate the Lagrangian multipliers solutions. To this end, this section proposes a nested SAGP algorithm to solve the EC-CMSTA model which is presented as follows.

Step 0 (initiation). Set parameters of SAGP within the given range: $\delta \in (0, 1)$, $\mu \in [0.5, 1]$, $\varepsilon > 0$, $\alpha_{\max} > 0$, $\alpha^{(0)} > 0$, $\gamma^{(0)} = \alpha^{(0)}$. Set the iteration counter: $n = 0$, and the initial solution which should be a feasible solution: $\mathbf{u}^{(0)} = (\mathbf{u}_a^{(0)} = 0, a \in E)^T$.

Step 1 (self-adaptive scaling procedure). This step aims to find the largest step size $\alpha^{(n+1)}$ that satisfies the given criteria in (27). To this end, let $\alpha^{(n+1)} = \mu^{l^{(n)}} \gamma^{(n)}$, and set nonnegative integer $l^{(n)} = 0$. We perform the following procedures, if the given criterion is satisfied, terminate and go to Step 2; otherwise, let $l^{(n)} = l^{(n)} + 1$, and perform the following procedures again.

Step 1.1. Given vector $\mathbf{u}^{(n)}$, evaluate the CMSTA with generalized travel time functions (23), using SAGP algorithm, and obtain the equilibrium flows $(\mathbf{v}^{(n)}, \mathbf{q}^{(n)})$. The solutions are subsequently used to evaluate the environmental measurement function $\mathbf{g}(\mathbf{u}^{(n)})$.

Step 1.2. Perform the projection operation for $\mathbf{u}^{(n)}$ to obtain a tentative vector $\bar{\mathbf{u}}^{(n+1)}$, i.e., $\bar{\mathbf{u}}^{(n+1)} = P_{\Omega}[\mathbf{u}^{(n)} - \alpha^{(n+1)} \cdot \mathbf{g}(\mathbf{u}^{(n)})]$.

Step 1.3. Let $\bar{\mathbf{u}}^{(n+1)} = \max[0, \mathbf{u}^{(n)} - \alpha^{(n)} \mathbf{g}(\mathbf{u}^{(n)})]$. If the following inequality is satisfied

$$\begin{aligned} & (2 - \delta) \alpha^{(n+1)} (\mathbf{u}^{(n)} - \bar{\mathbf{u}}^{(n+1)})^T (\mathbf{g}(\mathbf{u}^{(n)}) - \mathbf{g}(\bar{\mathbf{u}}^{(n+1)})) \\ & + (\alpha^{(n+1)})^2 \|\mathbf{g}(\mathbf{u}^{(n)}) - \mathbf{g}(\bar{\mathbf{u}}^{(n+1)})\|^2 \\ & \geq \max \left\{ 0, \frac{(\alpha^{(n+1)})^2 - (\alpha^{(n)})^2}{(\alpha^{(n)})^2} \|\mathbf{u}^{(n)} - \bar{\mathbf{u}}^{(n+1)}\|^2 \right\} \end{aligned} \quad (27)$$

go Step 2. Otherwise, let $l^{(n)} = l^{(n)} + 1$, update step size $\alpha^{(n+1)}$, and go back to Step 1.1.

Step 2 (selection of $\gamma^{(n+1)}$). If the following condition is satisfied, let $\gamma^{(n+1)} = \max\{\alpha^{(n+1)}/\mu, \alpha_{\max}\}$.

$$\begin{aligned} & 0.5 \alpha^{(n+1)} (\mathbf{u}^{(n)} - \bar{\mathbf{u}}^{(n+1)})^T (\mathbf{g}(\mathbf{u}^{(n)}) - \mathbf{g}(\bar{\mathbf{u}}^{(n+1)})) \\ & + (\alpha^{(n+1)})^2 \|\mathbf{g}(\mathbf{u}^{(n)}) - \mathbf{g}(\bar{\mathbf{u}}^{(n+1)})\|^2 \\ & \geq \max \left\{ 0, \frac{(\alpha^{(n+1)})^2 - (\alpha^{(n)})^2}{(\alpha^{(n)})^2} \|\mathbf{u}^{(n)} - \bar{\mathbf{u}}^{(n+1)}\|^2 \right\} \end{aligned} \quad (28)$$

Step 3 (convergence test). If the convergence criterion is satisfied, terminate; otherwise, let iteration counter: $n = n + 1$, go back to Step 1.

Note that the CMSTA with generalized travel time functions in Step 1.1 is also solved by the SAGP algorithm, which is not repeated here for conciseness.

6. Numerical Examples

In this section, two experiments based on a linear network and the Sioux-Falls network are provided to demonstrate the features of the proposed EC-CMSTA and validity of the proposed solution algorithm.

It is well known that carbon monoxide (CO), carbon dioxide (CO₂), sulphur dioxide (SO₂), nitrogen oxides (NO_x), respirable suspended particulates (RSP or PM₁₀), fine suspended particulates (FSP or PM_{2.5}), and volatile organic compounds (VOC) are the main types of vehicular emission pollutants that are critical for air quality protection. Among others, CO is taken as an important indicator of the level of vehicle atmosphere pollution, mainly because (1) vehicles contribute to the majority of CO emissions in the atmosphere; (2) CO is one of the most critical pollutants among the various types of vehicular emissions; (3) the emission functions of other pollutants are similar to that of CO [11]. In addition, this consideration has also been used in a number of recent studies, e.g., Yin and Lawphongpanich [37]; Nagurney, Qiang and Nagurney [38]; Chen, Zhou and Ryu [10]; Li et al. [11]; Xu, Chen and Cheng [1]; Yang, Ban and Ma [39]; Kitthamkesorn et al. [2]. Therefore, we use CO as an illustration to model vehicle emission. The nonlinear macroscopic formula of CO emission rate proposed by Wallace et al. [6] is used for analysis as shown below:

$$g_a(t_a) = 0.2038 \cdot t_a \cdot \exp\left(\frac{0.7962 \cdot L_a}{t_a}\right) \quad (29)$$

where $g_a(t_a)$ is the CO emission rate on link a measured in grams per hour. L_a is the length of link a measured in kilometers, and t_a is the link travel time measured in minutes. This nonlinear function can be shown to be convex and continuously differentiable. The Bureau of Public Road (BPR) type function is adopted for the link travel time, i.e., $t_a = t_a^0(1 + 0.15(v_a/C_a)^4)$ for road links, and $t_a = t_a^0(1 + 0.1(v_a/C_a)^4)$ for rail and RPR links. The dispersion parameter of MNL mode is set as $\theta = 1.0$.

6.1. A Linear Corridor. As shown in Figure 1, consider a linear corridor $G(N, A)$, with one rail line (dotted line) from suburb o to the urban city area d and another highway (solid line)

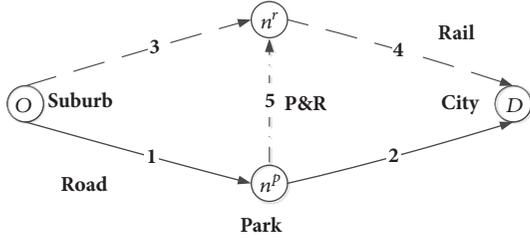


FIGURE 1: A linear corridor with RPR facility.

TABLE 1: Attributes of the linear corridor.

Link Type	Auto		Rail		RPR
Link No.	1	2	3	4	5
Free Flow Travel Time (t_a^0)	6	4	12	5	3
Capacity (C_a)	250	200	350	350	350

TABLE 2: Modal share and equilibrium travel time.

Modes	With RPR scheme	
	Demand	Travel time
Auto	250.45	16.66
Train	150.15	17.16
RPR	99.38	17.58

connecting (o, d) . The OD demand is assumed to be 500 per hour. One RPR site is constructed at n^p , adjacent to the train station n^r . Travellers transfer from the RPR site n^p to the train station n^r through the RPR link (link 5). Link 2 passes into an environmentally sensitive urban area, which exerts restriction on the total CO emissions on link 2. Using the CO emission rate function defined by (29), the nonlinear environmental constraint on link a can be expressed as

$$0.2038 \cdot t_a \cdot \exp\left(\frac{0.7962L_a}{t_a}\right) \cdot v_a - \bar{g}_a \leq 0, \quad \forall a \in A \quad (30)$$

where \bar{g}_a is the CO emission threshold on link a . Attributes of the network are given in Table 1.

Setting the environmental constraints at 500, we first examine the correctness of the results. As expected, the results in Table 2 satisfy the modal demand conservation condition and the MNL based modal choice condition. The Lagrangian multiplier on link 2 equals 1.73, permitting the emission on link 2 subject to the requirements.

We proceed to explore the influence of the environmental constraint on the flow shift of modal demands, link flows, and total system travel time by varying the emission constraint \bar{g}_2 between 0 and 800. The corresponding modal demand shift pattern under environmental constraint is given in Figure 2. We can see that as \bar{g}_2 increases (relax the CO emission constraints), the auto demand first increases and then becomes stable; meanwhile, the demands of rail and RPR show the crosscurrent. It makes sense because relaxing \bar{g}_2 is similar to increase the road capacity. When the emission level is beyond 631.47, constraint (30) is not blinding anymore which means the PRP facilities are able to achieve the environmental

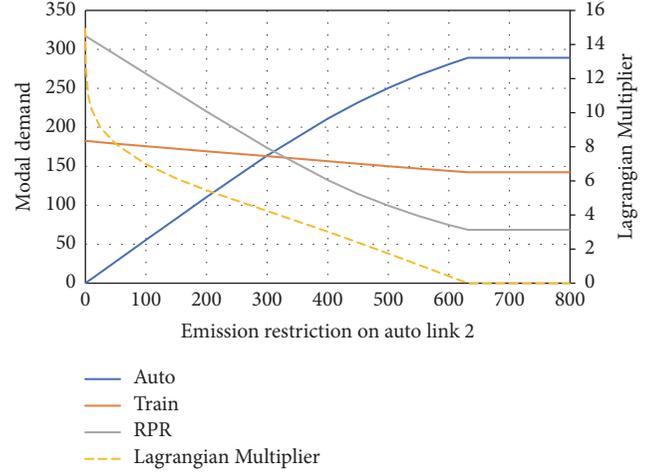


FIGURE 2: Modal split pattern with various environmental constraint level.

protection requirements without additional regulations or restrictions. The dotted line indicates the evolution of Lagrangian multiplier on link 2, u_2^* . As the environmental regulation can be strengthened, u_2^* keeps increasing and trends to be infinity, when $\bar{g}_2 = 0$.

Detailed results of link flows, link travel time, and Lagrangian multiplier are shown in Table 3. As marked by bold and italic text in Table 3 the minimal system travel cost is achieved when \bar{g}_2 is increased to around 631.47. In other words, setting $\bar{g}_2 \geq 631.47$ is equivalent to have no CO emission constraint on link 2, which reduces the problem into CMSTA. Meanwhile, the Lagrangian multiplier u_2^* reduces to zero as well indicating that there is no restrictions for vehicles which come into this area.

6.2. Sioux-Falls Network. Another numerical example built from the Sioux-Falls network (Figure 3) is used to further access the performance of the proposed model and solution algorithm for solving the environmental constrained CMSTA problem. This study network consists of 30 nodes and 76 links. The area highlighted by the red ellipse is the CBD district. It has four rail lines between the suburb areas and the city loop. Detail attributes of this network are tabulated in Appendix D. We focus on the travel demands in commuting hours, where the majority of travel demands origin from the suburbs and go into business districts. There are 24 OD pairs and the OD demand data are given in Appendix E. Four sites (nodes 3, 6, 19, and 14) which locate close to rail lines are selected as RPR car parks where RPR users park their car and transfer to rail mode.

The column generation scheme is adopted in this case and incorporated into our proposed solution algorithm to generate paths in evaluating the equilibrium flows. In the multimodal transport network, we first examine the effect of the proposed RPR scheme with the CMSTA model and access the emissions in the road subnetwork. 5 links in the CBD district are subsequently identified as the critical links, and be further grouped into the link set $\bar{A} = \{14, 22, 29, 46, 48\}$. As shown in Figure 4, the provision of RPR facilities significantly

TABLE 3: Link flows and link cost with varied restricted CO emission level.

\bar{g}_2	Link 1		Link 2		u^*	Link 3		Link 4		Link 5	
	Time	Flow	Time	Flow		Time	Flow	Time	Flow	Time	Flow
0	8.33	317.29	4.00	0.00	$+\infty$	12.09	182.71	7.08	499.94	3.20	317.23
100	8.55	324.24	4.00	55.32	6.99	12.08	175.76	6.30	444.68	3.10	268.92
200	8.76	330.69	4.06	110.34	5.45	12.07	169.31	5.77	389.66	3.05	220.35
300	8.97	337.09	4.27	163.54	4.24	12.06	162.91	5.43	336.46	3.02	173.55
400	9.21	343.56	4.75	211.36	3.02	12.05	156.44	5.23	288.64	3.01	132.20
500	9.45	349.85	5.48	250.47	1.73	12.04	150.15	5.13	249.53	3.00	99.38
600	9.69	355.77	6.34	281.02	0.41	12.03	144.23	5.08	218.98	3.00	74.75
631.47	9.77	357.57	6.62	289.22	0.00	12.03	142.43	5.07	210.78	3.00	68.35
700	9.77	357.57	6.62	289.22	0.00	12.03	142.43	5.07	210.78	3.00	68.35
800	9.77	357.57	6.62	289.22	0.00	12.03	142.43	5.07	210.78	3.00	68.35
900	9.77	357.57	6.62	289.22	0.00	12.03	142.43	5.07	210.78	3.00	68.35

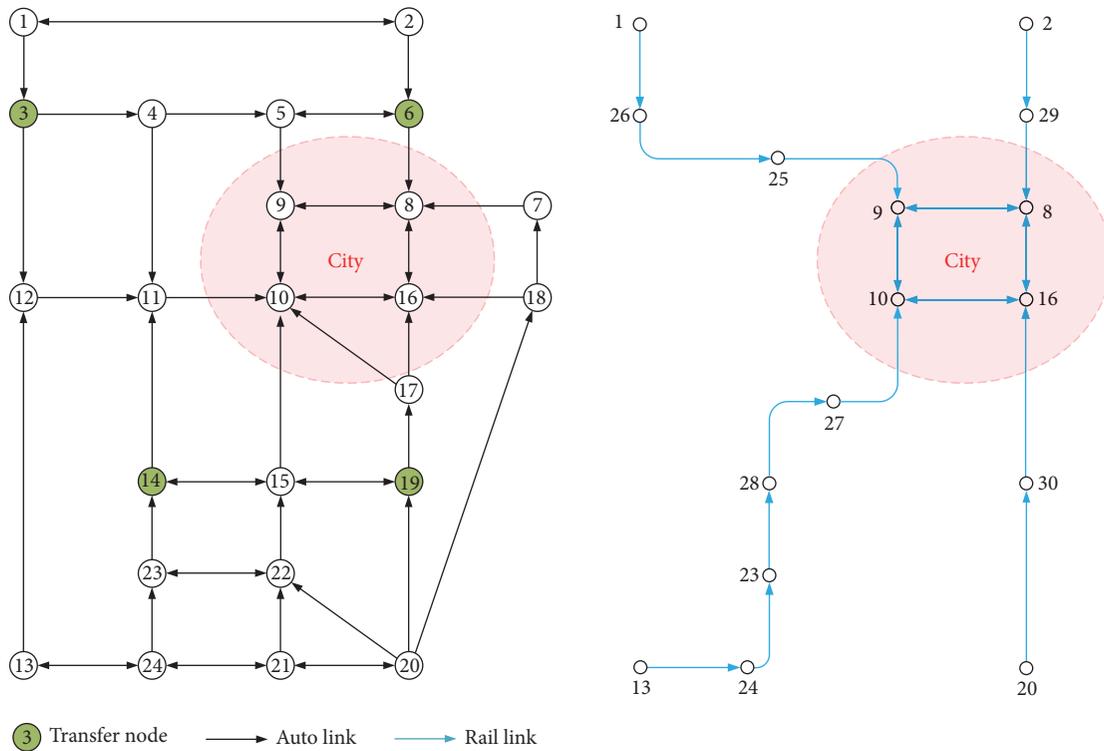


FIGURE 3: Modified Sioux-Falls network with RPR scheme.

influence travellers' mode choice as well as route choice behaviour. 14.03% of previous autodrivers and 5.51% of previous train riders shift to RPR mode, due to the utility provided by RPR services. The amount of emissions on critical links also significantly reduced with the provision of RPR schemes. This phenomenon indicates RPR schemes could alleviate the environmental deterioration by redistributing the flow and modal demand in the network and reflects the rationality of modelling users modal choice and route choice decisions in a combined model, i.e., CMSTA.

We proceed to exert environmental constraints on these critical links. Here, we consider five scenarios of the thresholds and summarize in Table 4. Parameters in the SAGP

algorithm are set as follows: $\mu = 0.85$, $\delta = 0.8$, and $\alpha^{(0)} = 1.0$. In order to reflect the convergence trend of the SAGP algorithm apparently, we take the following logarithmic value of the maximum absolute error as a performance index, namely,

$$\varepsilon = \log_{10} \left(\max \left(\|T_a - g_a v_a\|_2, \forall a \in \bar{A} \right) \right) \quad (31)$$

The SAGP algorithm is then used to solve the environmental constrained CMSTA in terms of the five scenarios provided in Table 4. Figure 5 depicts the logarithmic value of the maximum absolute error versus the number of iterations used for solving scenarios 1, 2, 3, and 4. It clearly shows that the

TABLE 4: Five scenarios of environmental constraints.

Link ID	Scenario 1 Threshold T_a^1	Scenario 2 Threshold T_a^2	Scenario 3 Threshold T_a^3	Scenario 4 Threshold T_a^4	Scenario 5 Threshold T_a^5	Emission	Link flow
14	500	600	700	800	900	724.16	320.28
22	500	600	700	800	900	1007.05	370.93
29	500	600	700	800	900	796.13	319.51
46	500	600	700	800	900	866.88	351.1
48	500	600	700	800 <td 900	735.61	531.39	

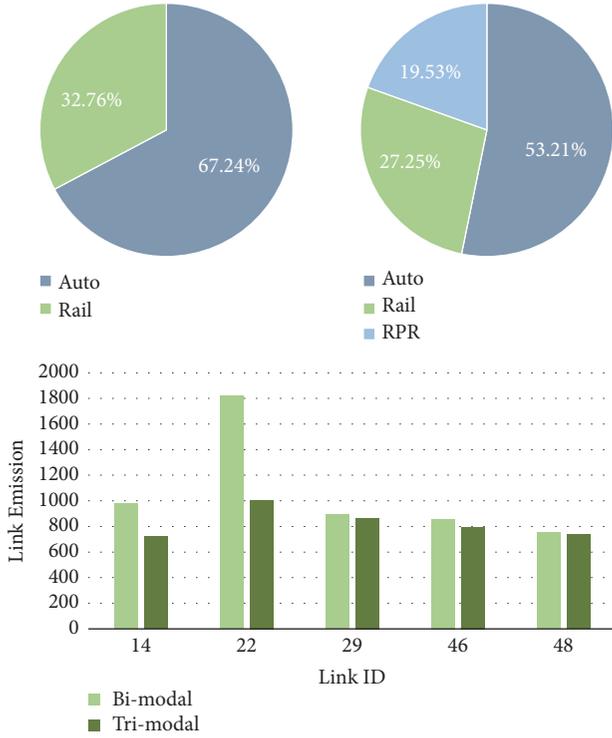


FIGURE 4: Mode share and emissions with/without RPR scheme.

algorithm can converge to a solution of VI model $VI(\mathbf{g}, \bar{\Omega})$. In scenario 5, the solution converges much faster than other scenarios because only link 22 is constrained.

Table 5 gives the resultant link emissions and corresponding Lagrangian multiplier for each scenario. We can observe that the emissions on each link are bounded by the required thresholds (some link emissions slightly exceed the requirements caused by the computation error). The corresponding multiplier takes positive value only on the links with emission bounded by the threshold. These two phenomena tally with (16)-(18), which numerically verify the effectiveness of the VI model $VI(\mathbf{g}, \bar{\Omega})$ and the SAGP algorithm. In addition, from scenario 1 to scenario 5, it can be seen that, by tightening the thresholds, the corresponding Lagrangian multipliers become larger, which indicates that it would be more costly to restrict the emission to a much lower level.

Table 6 further investigates the modal shift trend from scenario 1 to scenario 5. As the values of threshold getting

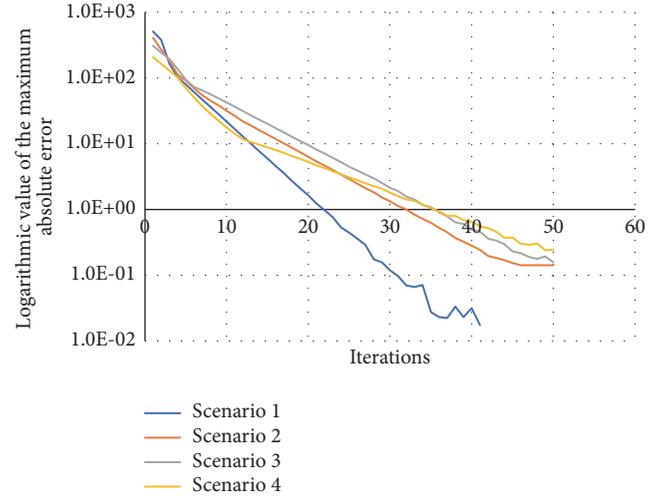


FIGURE 5: Convergence trend of the SAGP algorithm.

smaller, travellers who use automode become less. The environmental regulations in road network could significantly promote the modal shift from auto to “green” travel modes (i.e., Train and RPR). Taking scenario 1, for example, the automode share is depressed about 11.11%, from 53.21% to 42.10%, where 5.38% of travel demand shifts to rail mode and 5.72% rest shifts to RPR mode. This phenomenon indicates that environmental regulations significantly influence not only travellers’ route choice behaviour but also mode choice behaviour.

To further see the impacts of demand level on the Lagrangian multipliers, a sensitivity test is carried out on the dispersion parameter on the modal choice function (8). Without loss of generality, the first scenario of threshold setting is taken for this test, and then the value of θ is changed from 0.5 to 0.9 with an interval of 0.1. The Lagrangian multipliers obtained by the SAGP algorithm in these five cases are presented in Table 7. We noted that a larger value of θ results in a larger level of autodemand level and the road network congestion level is higher accordingly. Table 7 indicates the generalized CMSTA conditions can still be fulfilled. As the demand level ascends, the corresponding Lagrangian multipliers also increases, meaning that it is more difficult to maintain the threshold constraints at a higher demand level.

We further compare the network-level performance before and after implementing the set of environmental

TABLE 5: Link emission and corresponding Lagrangian multiplier.

Link ID	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Without
	Emission	LM									
14	499.99	1.64	600.06	0.93	700.07	0.22	728.78	0.00	726.51	0.00	724.16
22	500.00	2.29	600.03	1.74	700.10	1.20	800.17	0.64	900.00	0.38	1007.05
29	500.00	0.79	600.00	0.23	700.03	0.10	800.06	0.03	867.92	0.00	866.88
46	500.02	2.32	600.14	1.5	700.09	0.77	800.21	0.15	803.76	0.00	796.13
48	500.03	0.86	600.02	0.71	699.99	0.32	743.59	0.00	736.09	0.00	735.61

*LM denotes Lagrangian multiplier.

TABLE 6: Mode share with different threshold levels.

Mode	With Threshold					Without Threshold
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	
Auto	42.10%	45.68%	48.96%	51.53%	52.35%	53.21%
Train	32.64%	30.64%	28.89%	27.61%	27.39%	27.25%
RPR	25.26%	23.68%	22.15%	20.86%	20.27%	19.53%

TABLE 7: Sensitivity test on the dispersion parameter.

link ID	$\theta = 0.5$		$\theta = 0.6$		$\theta = 0.7$		$\theta = 0.8$		$\theta = 0.9$	
	Emission	LM								
14	499.99	1.64	500.00	1.70	499.98	1.73	500.01	1.74	500.02	1.75
22	500.00	2.29	500.00	2.02	500.00	1.82	500.02	1.66	500.01	1.54
29	500.00	0.79	500.01	0.68	500.01	0.59	500.00	0.52	500.01	0.48
46	500.02	2.32	500.03	2.38	500.03	2.41	500.00	2.42	500.06	2.43
48	500.03	0.86	500.01	0.85	499.94	0.84	500.02	0.84	499.96	0.83

*LM denotes Lagrangian multiplier.

restraints. The first scenario of threshold setting is still taken for this test. Without loss of generality, the total travel time (TTT) and total vehicle distances (TVD) travelled are used in this study to measure the network-wide performance. From Table 8 we can observe that implementation of environmental constraints decreases both TTT and TVD (-13.89% decrease of TTT and -13.19% decrease of TVM), which implies that some travellers must be shifting their modes in order to benefit the air quality of the CBD area.

7. Conclusions

Environmental protection is an important consideration in planning a transportation system. Environmental constraint is a useful means to explicitly reflect environmental protection requirements. In this paper, we studied a combined environmental protection strategy in the multimodal transport network: (1) a rail-based park-and-ride scheme to assist mode shift to reduce the overall vehicle travel demands in the road network and (2) an area-based regulation scheme to deal with the environmental deterioration caused by heterogeneity of traffic flows. To quantitatively evaluate the effectiveness of the proposed scheme, we built an environmental constrained combined modal split and traffic assignment (EC-CMSTA) model. Specifically, we developed an equivalent VI reformulation for the EC-CMSTA. Different from

the explicit formulation as a VI or NCP, the proposed reformulation adapts the concept of fix-point to reformulate the CMSTA conditions, which is, later, incorporated into the VI formulated nonlinear complementarity conditions associated with the nonlinear environmental constraints. The proposed VI reformulation can handle a general nonlinear environmental constraint structure and a general link and route cost structure, enhancing the modelling adaptability and flexibility, and in addition permits a number of efficient algorithms for its solution. A customized projection algorithm with self-adaptive step size scheme is adopted in this paper. Numerical examples demonstrated the EC-CMSTA model has the potential to enhance the behavioural modelling of network users' travel decisions and to assist in quantitatively evaluating the effectiveness of RPR schemes and environmental regulations.

Several directions are worthy of further investigation. First, in this paper, we use CO as an indicator to model vehicle emission. Further study will take other types of emission into consideration, e.g., the CO₂. Second, a number of efficient solution algorithms proposed by recent studies, e.g., Ryu, Chen, and Choi [34]; Ryu, Chen, and Choi [40], should be explored to further improve the solution efficiency. Third, we plan to test our model on larger networks with more real world applications. For very large-scale problems, existing algorithms may be not efficient and specific algorithms should be developed.

TABLE 8: Network performance with and without environmental regulations.

Network performance	Without environmental Regulation (a)	With environmental regulation (b)	Relative difference [(b-a)/a]
Total travel time	41442.50	35685.36	-13.89%
Total vehicle distance	38995.10	33851.32	-13.19%

Appendix

A. Proof of Proposition 5

Vector function $\mathbf{g}(\mathbf{u})$ is strictly monotone on $\bar{\Omega}$; namely,

$$(\mathbf{g}(\mathbf{u}') - \mathbf{g}(\mathbf{u}''))^T (\mathbf{u}' - \mathbf{u}'') > 0, \quad \forall \mathbf{u}', \mathbf{u}'' \in \bar{\Omega} \quad (\text{A.1})$$

Proof. For any two distinct nonnegative Lagrangian multiplier vectors \mathbf{u}' and \mathbf{u}'' , let $\mathbf{v}(\mathbf{u}')$ and $\mathbf{v}(\mathbf{u}'')$ denote the corresponding CMSTA link flow solutions, and similarly, let $\mathbf{f}(\mathbf{u}')$, $\mathbf{f}(\mathbf{u}'')$, $\mathbf{q}(\mathbf{u}')$, and $\mathbf{q}(\mathbf{u}'')$ denote the corresponding CMSTA path flow and modal demand solutions. We have

$$\mathbf{v}(\mathbf{u}') = \Delta \mathbf{f}(\mathbf{u}') \quad (\text{A.2})$$

$$\mathbf{v}(\mathbf{u}'') = \Delta \mathbf{f}(\mathbf{u}'') \quad (\text{A.3})$$

$$\mathbf{f}(\mathbf{u}') = \mathbf{q}(\boldsymbol{\tau}')^T \cdot \mathbf{p}(\hat{\mathbf{c}}') \quad (\text{A.4})$$

$$\mathbf{f}(\mathbf{u}'') = \mathbf{q}(\boldsymbol{\tau}'')^T \cdot \mathbf{p}(\hat{\mathbf{c}}'') \quad (\text{A.5})$$

$$\mathbf{q}(\hat{\boldsymbol{\tau}}') = \bar{\mathbf{q}}^T \cdot \bar{\mathbf{p}}(\hat{\boldsymbol{\tau}}') \quad (\text{A.6})$$

$$\mathbf{q}(\hat{\boldsymbol{\tau}}'') = \bar{\mathbf{q}}^T \cdot \bar{\mathbf{p}}(\hat{\boldsymbol{\tau}}'') \quad (\text{A.7})$$

where Δ defines the link-path incidence, $\hat{\mathbf{c}}' = \Delta^T \hat{\mathbf{t}}(\mathbf{v}(\mathbf{u}'), \mathbf{u}')$, $\hat{\mathbf{c}}'' = \Delta^T \hat{\mathbf{t}}(\mathbf{v}(\mathbf{u}''), \mathbf{u}'')$, $\mathbf{p}(\hat{\mathbf{c}}') = (p_k^{m,od}(\hat{\mathbf{c}}'), \forall k \in K_m^{od}, m \in M_{od}, od \in W)^T$, $\mathbf{p}(\hat{\mathbf{c}}'') = (p_k^{m,od}(\hat{\mathbf{c}}''), \forall k \in K_m^{od}, m \in M_{od}, od \in W)^T$, $\hat{\boldsymbol{\tau}}' = (\tau_m^{od}(\hat{\mathbf{c}}'), \forall m \in M_{od}, od \in W)^T$, $\hat{\boldsymbol{\tau}}'' = (\tau_m^{od}(\hat{\mathbf{c}}''), \forall m \in M_{od}, od \in W)^T$, $\bar{\mathbf{p}}(\hat{\boldsymbol{\tau}}') = (p_m^{od}(\hat{\boldsymbol{\tau}}'), \forall m \in M_{od}, od \in W)^T$, $\bar{\mathbf{p}}(\hat{\boldsymbol{\tau}}'') = (p_m^{od}(\hat{\boldsymbol{\tau}}''), \forall m \in M_{od}, od \in W)^T$, and $\bar{\mathbf{q}} = (\bar{q}^{od}, \forall od \in W)^T$.

According to the definition of CMSTA conditions, the modal travel time τ_m^{od} is taken as the equilibrium path travel time of mode m , it follows that

$$\tau_m^{od} \leq c_k^{m,od}, \quad \forall k \in K_m^{od}; \quad (\text{A.8})$$

together with the flow conservation condition (5), we have

$$\mathbf{q}(\boldsymbol{\tau}')^T \boldsymbol{\tau}' \leq \mathbf{f}(\mathbf{u}')^T \mathbf{c}' \quad (\text{A.9})$$

$$\mathbf{q}(\boldsymbol{\tau}'')^T \boldsymbol{\tau}'' \leq \mathbf{f}(\mathbf{u}'')^T \mathbf{c}'' \quad (\text{A.10})$$

Subtracting (A.9) and (A.10) yields that

$$\mathbf{q}(\boldsymbol{\tau}')^T (\boldsymbol{\tau}' - \boldsymbol{\tau}'') \leq \mathbf{f}(\mathbf{u}')^T (\mathbf{c}' - \mathbf{c}'') \quad (\text{A.11})$$

$$\mathbf{q}(\boldsymbol{\tau}'')^T (\boldsymbol{\tau}' - \boldsymbol{\tau}'') \leq \mathbf{f}(\mathbf{u}'')^T (\mathbf{c}' - \mathbf{c}''); \quad (\text{A.12})$$

namely,

$$\begin{aligned} & (\mathbf{q}(\boldsymbol{\tau}') - \mathbf{q}(\boldsymbol{\tau}''))^T (\boldsymbol{\tau}' - \boldsymbol{\tau}'') \\ & \leq (\mathbf{f}(\mathbf{u}') - \mathbf{f}(\mathbf{u}''))^T (\mathbf{c}' - \mathbf{c}'') \end{aligned} \quad (\text{A.13})$$

The monotonicity of function (8) implies that

$$(\mathbf{q}(\boldsymbol{\tau}') - \mathbf{q}(\boldsymbol{\tau}''))^T (\boldsymbol{\tau}' - \boldsymbol{\tau}'') > 0 \quad (\text{A.14})$$

According to (A.13) and (A.14), we thus obtain

$$(\mathbf{f}(\mathbf{u}') - \mathbf{f}(\mathbf{u}''))^T (\mathbf{c}' - \mathbf{c}'') > 0 \quad (\text{A.15})$$

As $\mathbf{c}' = \Delta^T \hat{\mathbf{t}}(\mathbf{v}(\mathbf{u}'), \mathbf{u}')$ and $\mathbf{c}'' = \Delta^T \hat{\mathbf{t}}(\mathbf{v}(\mathbf{u}''), \mathbf{u}'')$, (A.15) can be equivalently written as

$$\begin{aligned} & (\mathbf{f}(\mathbf{u}') - \mathbf{f}(\mathbf{u}''))^T \\ & \cdot (\Delta^T \hat{\mathbf{t}}(\mathbf{v}(\mathbf{u}'), \mathbf{u}') - \Delta^T \hat{\mathbf{t}}(\mathbf{v}(\mathbf{u}''), \mathbf{u}'')) > 0 \end{aligned} \quad (\text{A.16})$$

Rearranging the left side of (A.16), it follows that

$$\begin{aligned} & (\mathbf{v}(\mathbf{u}') - \mathbf{v}(\mathbf{u}''))^T \\ & \cdot (\Delta^T \hat{\mathbf{t}}(\mathbf{v}(\mathbf{u}'), \mathbf{u}') - \Delta^T \hat{\mathbf{t}}(\mathbf{v}(\mathbf{u}''), \mathbf{u}'')) > 0 \end{aligned} \quad (\text{A.17})$$

According to the definition of generalized link travel time function, (20), we have

$$(\mathbf{v}(\mathbf{u}') - \mathbf{v}(\mathbf{u}''))^T (\mathbf{u}' - \mathbf{u}'') > 0 \quad (\text{A.18})$$

Together with Assumption 1, $\mathbf{g}(\mathbf{u})$ is a strictly monotone function with respect to link flow solutions $\mathbf{v}(\mathbf{u})$; it follows that

$$(\mathbf{g}(\mathbf{u}') - \mathbf{g}(\mathbf{u}''))^T (\mathbf{u}' - \mathbf{u}'') > 0 \quad (\text{A.19})$$

which implies $\mathbf{g}(\mathbf{u})$ is strictly monotone with respect to \mathbf{u} on $\bar{\Omega}$. \square

B. Proof of Proposition 6

Vector function $\mathbf{g}(\mathbf{u})$ is Lipschitz continuous on $\bar{\Omega}$; namely,

$$\|\mathbf{g}(\mathbf{u}') - \mathbf{g}(\mathbf{u}'')\|_2 \leq L \|\mathbf{u}' - \mathbf{u}''\|, \quad \forall \mathbf{u}', \mathbf{u}'' \in \bar{\Omega} \quad (\text{B.1})$$

Proof. Following the approach presented by Meng, Liu, and Wang [28], it is easily to conclude that $\mathbf{v}(\mathbf{u})$ is continuously differentiable in a neighbourhood of \mathbf{u}_0 according to the implicit function theorem [41].

Since $\mathbf{v}(\mathbf{u})$ is continuously differentiable on $\bar{\Omega}$, its Jacobian matrix $\nabla_{\mathbf{u}} \mathbf{v}(\mathbf{u})$ is thus continuous on $\bar{\Omega}$. The two-norm of $\nabla_{\mathbf{u}} \mathbf{v}(\mathbf{u})$ is therefore bounded from the nonempty closed and convex set $\bar{\Omega}$; namely, there is a positive constant such that

$$\|\nabla_{\mathbf{u}} \mathbf{v}(\mathbf{u})\|_2 \leq L_1, \quad \forall \mathbf{u} \in \bar{\Omega} \quad (\text{B.2})$$

According to the Theorem 3.2.4 of Ortega and Rheinboldt [41] (mean-value theorem), it can be seen that

$$\|\mathbf{v}(\mathbf{u}') - \mathbf{v}(\mathbf{u}'')\|_2 \leq L_1 \|\mathbf{u}' - \mathbf{u}''\|_2, \quad \forall \mathbf{u}', \mathbf{u}'' \in \bar{\Omega} \quad (\text{B.3})$$

Similarly, according to **Assumption 1**, $\mathbf{g}(\mathbf{v})$ is a continuous differentiable function w.r.t. \mathbf{v} . The Jacobian matrix $\nabla_{\mathbf{v}}\mathbf{g}(\mathbf{v})$ is thus continuous on Ω which is also nonempty closed and convex. Thus, there is a positive constant such that

$$\|\nabla_{\mathbf{v}}\mathbf{g}(\mathbf{v})\|_2 \leq L_2, \quad \forall \mathbf{v} \in \Omega \quad (\text{B.4})$$

and we have

$$\|\mathbf{g}(\mathbf{v}') - \mathbf{g}(\mathbf{v}'')\|_2 \leq L_2 \|\mathbf{v}' - \mathbf{v}''\|_2, \quad \forall \mathbf{v}', \mathbf{v}'' \in \Omega \quad (\text{B.5})$$

According to (B.3) and (B.5), we thus have

$$\begin{aligned} \|\mathbf{g}(\mathbf{u}') - \mathbf{g}(\mathbf{u}'')\|_2 &\leq L_2 \|\mathbf{v}' - \mathbf{v}''\|_2 \\ &\leq L_2 \cdot L_1 \|\mathbf{u}' - \mathbf{u}''\|_2, \end{aligned} \quad (\text{B.6})$$

$$\forall \mathbf{u}', \mathbf{u}'' \in \bar{\Omega}$$

This completes the proof. \square

C. Proof of Proposition 7

$\mathbf{u}^* = (u_a^*, \forall a \in E)$ is a solution of $VI(\mathbf{g}, \mathfrak{R}_+^{|\mathcal{E}|})$ if and only if \mathbf{u}^* and $\mathbf{v}(\mathbf{u}^*)$ fulfil the generalized CMSTA conditions (15)-(18).

Proof.

Necessary Condition. Suppose \mathbf{u}^* is a solution of $VI(\mathbf{g}, \mathfrak{R}_+^{|\mathcal{E}|})$, and we now show that \mathbf{u}^* fulfil the generalized CMSTA conditions.

We first demonstrate that $u_a^* < \bar{M}, \forall a \in E$. Assume that there is at least on element $b \in E$ with $u_b^* = \bar{M}$. Then, we define a feasible vector $\mathbf{u}' \in \mathfrak{R}_+^{|\mathcal{E}|}$.

$$\mathbf{u}' = (u'_a = u_a^*, \forall a \in E \setminus \{b\}, u'_b = 0.5\bar{M})^T. \quad (\text{C.1})$$

Substituting \mathbf{u} in the VI model $VI(\mathbf{g}, \mathfrak{R}_+^{|\mathcal{E}|})$ with vector \mathbf{u}' , it follows that

$$(-g_b(\mathbf{v}(\mathbf{u}')) \cdot (0.5\bar{M} - \bar{M})) \geq 0; \quad (\text{C.2})$$

namely,

$$g_b(\mathbf{v}(\mathbf{u}')) \geq 0 \quad (\text{C.3})$$

Now, we proceed to demonstrate $u_a^* < \bar{M}, \forall a \in E$ by using an apagogical approach. Note that, according to the monotony increasing of $\mathbf{g}(\mathbf{v})$, there exists a upper bound of link flow \bar{v}_b that, for any flow $v_b < \bar{v}_b$, we have

$$g_b(\mathbf{v}(\mathbf{u}^*)) < 0 \quad (\text{C.4})$$

Then, we proceed to show $v_a(\mathbf{u}^*) < \bar{v}_a, \forall a \in E$. Suppose at least one flow loaded path on the road network uses link b . According to the DUE conditions, for any flow loaded paths, they have identical path travel time. Denote one of these paths as k_o and its $\lambda_{k_o}^{m,od}(\mathbf{u}^*)$ follows that

$$\lambda_{k_o}^{m,od}(\mathbf{u}^*) = \sum_{a \in E} u_a^* \zeta_{e,k_o}^{m,od} > \bar{M} \quad (\text{C.5})$$

Therefore, its generalized path travel time and the corresponding modal travel time fulfil the condition:

$$c_{k_o}^{m,od}(\mathbf{u}^*) + \lambda_{k_o}^{m,od}(\mathbf{u}^*) > \bar{M} \quad (\text{C.6})$$

$$c_m^{od}(\mathbf{u}^*) + \lambda_m^{od}(\mathbf{u}^*) \geq c_{k_o}^{m,od}(\mathbf{u}^*) + \lambda_{k_o}^{m,od}(\mathbf{u}^*) > \bar{M} \quad (\text{C.7})$$

According to the monotonicity of function (8), we have

$$q_m^{od}(c_m^{od}(\mathbf{u}^*) + \lambda_m^{od}(\mathbf{u}^*)) \leq q_m^{od}(\bar{M}) < \frac{\bar{v}_b}{|W| \cdot |M|} \quad (\text{C.8})$$

In other words,

$$v_b < \sum_{od \in W} \sum_{m \in M_{od}} q_m^{od} < \bar{v}_b \quad (\text{C.9})$$

$$g_b(v_b(\mathbf{u}^*)) < 0 \quad (\text{C.10})$$

However, (C.10) contradicts (C.3), then the hypothesis $u_b^* = \bar{M}$ is incorrect. In other words,

$$u_a^* < \bar{M}, \quad \forall a \in E \quad (\text{C.11})$$

To further prove the necessary conditions, we construct the following two specific vectors for any link $a \in E$:

$$\mathbf{u}' = (u'_a = (1 - \gamma_1)u_a^* + \gamma_1\bar{M}, \forall a \in E)^T \quad (\text{C.12})$$

$$\mathbf{u}'' = (u''_a = \gamma_1 u_a^*, \forall a \in E)^T \quad (\text{C.13})$$

where $\gamma_1 \in (0, 1)$ is a given parameter. Substituting vector \mathbf{u} of VI model $VI(\mathbf{g}, \mathfrak{R}_+^{|\mathcal{E}|})$ with these two particular vectors, respectively, yields that

$$g_a(\mathbf{v})(\gamma_1(\bar{M} - u_a^*)) \leq 0 \quad (\text{C.14})$$

$$g_a(\mathbf{v})((\gamma_1 - 1)u_a^*) \leq 0 \quad (\text{C.15})$$

Since $M - u_a^* > 0$, (C.14) implies that

$$g_a(\mathbf{v}) \leq 0, \quad \forall a \in E \quad (\text{C.16})$$

In accordance with the fact that $u_a^* \geq 0$, we thus have

$$(-g_a(\mathbf{v})) \cdot u_a^* \geq 0, \quad \forall a \in E \quad (\text{C.17})$$

Hence, (C.17) in conjunction with (C.15) implies that

$$(-g_a(\mathbf{v})) \cdot u_a^* = 0, \quad \forall a \in E \quad (\text{C.18})$$

In other words, \mathbf{u}^* and $\mathbf{v}(\mathbf{u}^*)$ fulfil the generalized CMSTA conditions.

Sufficient Condition. Suppose \mathbf{u}^* fulfil the generalized CMSTA conditions (15)-(18); we have the following conditions:

$$-\mathbf{g}(\mathbf{v}(\mathbf{u}^*)) \geq \mathbf{0} \quad (\text{C.19})$$

$$\mathbf{g}(\mathbf{v}(\mathbf{u}^*))^T \cdot \mathbf{u}^* = 0 \quad (\text{C.20})$$

In addition, we have $\mathbf{u} \geq 0$. Therefore, according to (C.19) and (C.20), it follows that

$$\mathbf{g}(\mathbf{v}(\mathbf{u}^*))^T \cdot \mathbf{u} \geq \mathbf{g}(\mathbf{v}(\mathbf{u}^*))^T \cdot \mathbf{u}^* \quad (\text{C.21})$$

$$\mathbf{g}(\mathbf{v}(\mathbf{u}^*))^T \cdot (\mathbf{u} - \mathbf{u}^*) \geq 0 \quad (\text{C.22})$$

This completes the proof. \square

TABLE 9

ID	T	H	Fft	Capa	Type
1	8	9	3	400	Auto
2	9	8	3	400	Auto
3	9	10	3	400	Auto
4	10	9	3	400	Auto
5	10	16	3	400	Auto
6	16	10	3	400	Auto
7	8	16	3	400	Auto
8	16	8	3	400	Auto
9	1	3	3	800	Auto
10	3	4	4	1200	Auto
11	4	3	4	1200	Auto
12	4	5	5	800	Auto
13	5	4	5	800	Auto
14	5	9	5	800	Auto
15	1	2	8	600	Auto
16	3	12	6	1200	Auto
17	4	11	7	600	Auto
18	5	6	3	800	Auto
19	6	5	3	800	Auto
20	12	11	5	600	Auto
21	2	6	3	800	Auto
22	6	8	6	800	Auto
23	2	1	8	600	Auto
24	13	24	3	600	Auto
25	24	23	4	600	Auto
26	23	14	4	600	Auto
27	14	15	3	1200	Auto
28	15	14	3	800	Auto
29	15	10	6	800	Auto
30	13	12	7	800	Auto
31	24	21	3	600	Auto
32	21	24	3	600	Auto
33	23	22	3	600	Auto
34	22	23	3	600	Auto
35	21	22	4	600	Auto
36	22	15	4	800	Auto
37	21	20	3	600	Auto
38	20	19	5	600	Auto
39	19	17	3	1200	Auto
40	17	16	3	600	Auto
41	17	10	5	400	Auto
42	20	22	3	400	Auto
43	15	19	3	600	Auto
44	7	8	3	400	Auto
45	19	15	3	600	Auto
46	11	10	5	600	Auto
47	14	11	4	400	Auto
48	18	16	3	600	Auto
49	18	7	4	600	Auto
50	20	21	3	600	Auto
51	20	18	7	600	Auto
52	24	13	3	600	Auto
53	9	10	3	1500	Train
54	9	8	3	1500	Train
55	10	9	3	1500	Train
56	10	16	3	1500	Train

TABLE 9: Continued.

ID	T	H	Fft	Capa	Type
57	8	9	3	1500	Train
58	8	16	3	1500	Train
59	16	10	3	1500	Train
60	16	8	3	1500	Train
61	20	30	7	1000	Train
62	30	16	6	1000	Train
63	13	24	2	1000	Train
64	24	23	4	1000	Train
65	23	28	4	1000	Train
66	28	27	4	1000	Train
67	27	10	4	1000	Train
68	2	29	10	1000	Train
69	29	8	2	1000	Train
70	1	26	3	1000	Train
71	26	25	9	1000	Train
72	25	9	7	1000	Train
73	3	26	4	1000	RPR
74	6	29	4	1000	RPR
75	14	28	3	1000	RPR
76	19	30	3	1000	RPR

*T is tail node; H is head node; Fft denotes free flow time; Capa denotes capacity.

TABLE 10

ID	Origin	Destination	Demand
1	1	8	220
2	1	9	180
3	1	10	200
4	1	16	220
5	2	8	240
6	2	9	190
7	2	10	180
8	2	16	180
9	13	8	160
10	13	9	130
11	13	10	170
12	13	16	130
13	24	8	140
14	24	9	150
15	24	10	140
16	24	16	100
17	23	8	130
18	23	9	150
19	23	10	150
20	23	16	170
21	20	8	180
22	20	9	160
23	20	10	140
24	20	16	170

D. Attributes of Study Network

See Table 9.

E. OD Travel Demand of Study Network

See Table 10.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] X. Xu, A. Chen, and L. Cheng, "Reformulating environmentally constrained traffic equilibrium via a smooth gap function," *International Journal of Sustainable Transportation*, vol. 9, no. 6, pp. 419–430, 2015.
- [2] S. Kitthamkesorn, A. Chen, X. Xu, and S. Ryu, "Modeling mode and route similarities in network equilibrium problem with go-green modes," *Networks and Spatial Economics*, vol. 16, no. 1, pp. 33–60, 2016.
- [3] W. Y. Szeto, X. Jaber, and S. C. Wong, "Road network equilibrium approaches to environmental sustainability," *Transport Reviews*, vol. 32, no. 4, pp. 491–518, 2012.
- [4] X. Chen, Z. Liu, I. Kim, and J. Leng, "Optimal Sitting and Sizing of the Remote Park-and-Ride Scheme in the Multimodal Transport Network," in *Proceedings of the Transportation Research Board 97th Annual Meeting*, Washington DC, USA, 2018.
- [5] Z. Liu and Q. Meng, "Bus-based park-and-ride system: a stochastic model on multimodal network with congestion pricing schemes," *International Journal of Systems Science*, vol. 45, no. 5, pp. 994–1006, 2014.
- [6] C. E. Wallace, K. Courage, D. Reaves, Schoene. G., and G. Euler, TRANSYT-7F user's manual. In.
- [7] L. R. Rilett and C. M. Benedek, "Traffic assignment under environmental and equity objectives," *Transportation Research Record*, no. 1443, pp. 92–99, 1994.
- [8] C. M. Benedek and L. R. Rilett, "Equitable traffic assignment with environmental cost functions," *Journal of Transportation Engineering*, vol. 124, no. 1, pp. 16–22, 1998.
- [9] S. Sugawara and D. Niemeier, "How much can vehicle emissions be reduced? Exploratory analysis of an upper boundary using an emissions-optimized trip assignment," *Transportation Research Record Journal of the Transportation Research Board*, no. 1815, pp. 29–37, 2002.
- [10] A. Chen, Z. Zhou, and S. Ryu, "Modeling physical and environmental side constraints in traffic equilibrium problem," *International Journal of Sustainable Transportation*, vol. 5, no. 3, pp. 172–197, 2011.
- [11] Z.-C. Li, W. H. K. Lam, S. C. Wong, and A. Sumalee, "Environmentally sustainable toll design for congested road networks with uncertain demand," *International Journal of Sustainable Transportation*, vol. 6, no. 3, pp. 127–155, 2012.
- [12] A. Nagurney, P. Ramanujam, and K. K. Dhanda, "A multimodal traffic network equilibrium model with emission pollution permits: complicity vs noncompliance," *Transportation Research Part D: Transport and Environment*, vol. 3, no. 5, pp. 349–374, 1998.
- [13] E. Fernandez, J. de Cea, M. Florian, and E. Cabrera, "Network equilibrium models with combined modes," *Transportation Science*, vol. 28, no. 3, pp. 182–192, 1994.
- [14] R. García and A. Marín, "Network equilibrium with combined modes: models and solution algorithms," *Transportation Research Part B: Methodological*, vol. 39, no. 3, pp. 223–254, 2005.
- [15] Z.-C. Li, W. H. K. Lam, S. C. Wong, D.-L. Zhu, and H.-J. Huang, "Modeling park-and-ride services in a multimodal transport network with elastic demand," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1994, pp. 101–109, 2007.
- [16] W. H. Lam, Z. Li, S. C. Wong, and D. Zhu, "Modeling an elastic-demand bimodal transport network with park-and-ride trips," *Tsinghua Science and Technology*, vol. 12, no. 2, pp. 158–166, 2007.
- [17] C. Pineda, C. E. Cortés, P. Jara-Moroni, and E. Moreno, "Integrated traffic-transit stochastic equilibrium model with park-and-ride facilities," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 86–107, 2016.
- [18] Z. Song, Y. He, and L. Zhang, "Integrated planning of park-and-ride facilities and transit service," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 182–195, 2017.
- [19] S. Kitthamkesorn and A. Chen, "Alternate weibit-based model for assessing green transport systems with combined mode and route travel choices," *Transportation Research Part B: Methodological*, vol. 103, pp. 291–310, 2017.
- [20] Z. Liu, X. Chen, Q. Meng, and I. Kim, "Remote park-and-ride network equilibrium model and its applications," *Transportation Research Part B: Methodological*, vol. 117, pp. 37–62, 2018.
- [21] S. Wang, Q. Meng, and H. Yang, "Global optimization methods for the discrete network design problem," *Transportation Research Part B: Methodological*, vol. 50, pp. 42–60, 2013.
- [22] Z. Liu, S. Wang, and Q. Meng, "Optimal joint distance and time toll for cordon-based congestion pricing," *Transportation Research Part B: Methodological*, vol. 69, pp. 81–97, 2014.
- [23] T. Liu, H. Huang, H. Yang, and X. Zhang, "Continuum modeling of park-and-ride services in a linear monocentric city with deterministic mode choice," *Transportation Research Part B: Methodological*, vol. 43, no. 6, pp. 692–707, 2009.
- [24] D. Z. W. Wang and B. Du, "Reliability-based modeling of park-and-ride service on linear travel corridor," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2333, pp. 16–26, 2013.
- [25] B. Du and D. Z. W. Wang, "Continuum modeling of park-and-ride services considering travel time reliability and heterogeneous commuters – A linear complementarity system approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 71, no. 0, pp. 58–81, 2014.
- [26] Z. Liu, Q. Meng, and S. Wang, "Variational inequality model for cordon-based congestion pricing under side constrained stochastic user equilibrium conditions," *Transportmetrica A: Transport Science*, vol. 10, no. 8, pp. 693–704, 2014.
- [27] T. Larsson and M. Patriksson, "Side constrained traffic equilibrium models— analysis, computation and applications," *Transportation Research Part B: Methodological*, vol. 33, no. 4, pp. 233–264, 1999.
- [28] Q. Meng, Z. Liu, and S. Wang, "Asymmetric stochastic user equilibrium problem with elastic demand and link capacity constraints," *Transportmetrica A: Transport Science*, vol. 10, no. 4, pp. 304–326, 2014.
- [29] C. F. Daganzo, "Stochastic network equilibrium with multiple vehicle types and asymmetric, indefinite link cost Jacobians," *Transportation Science*, vol. 17, no. 3, pp. 282–300, 1983.
- [30] G. E. Cantarella, "A general fixed-point approach to multimode multi-user equilibrium assignment with elastic demand," *Transportation Science*, vol. 31, no. 2, pp. 107–128, 1997.
- [31] A. Chen, X. Xu, S. Ryu, and Z. Zhou, "A self-adaptive Armijo stepsize strategy with application to traffic assignment models

- and algorithms,” *Transportmetrica A: Transport Science*, vol. 9, no. 8, pp. 695–712, 2011.
- [32] A. Chen, Z. Zhou, and X. Xu, “A self-adaptive gradient projection algorithm for the nonadditive traffic equilibrium problem,” *Computers & Operations Research*, vol. 39, no. 2, pp. 127–138, 2012.
- [33] Z. Zhou, A. Chen, and S. C. Wong, “Alternative formulations of a combined trip generation, trip distribution, modal split, and trip assignment model,” *European Journal of Operational Research*, vol. 198, no. 1, pp. 129–138, 2009.
- [34] S. Ryu, A. Chen, and K. Choi, “Solving the combined modal split and traffic assignment problem with two types of transit impedance function,” *European Journal of Operational Research*, vol. 257, no. 3, pp. 870–880, 2017.
- [35] B. S. He, H. Yang, Q. Meng, and D. R. Han, “Modified Goldstein-Levitin-Polyak projection method for asymmetric strongly monotone variational inequalities,” *Journal of Optimization Theory and Applications*, vol. 112, no. 1, pp. 129–143, 2002.
- [36] D. Han and W. Sun, “A new modified Goldstein-Levitin-Polyak projection method for variational inequality problems,” *Computers Mathematics with Applications*, vol. 47, no. 12, pp. 1817–1825, 2004.
- [37] Y. Yin and S. Lawphongpanich, “Internalizing emission externality on road networks,” *Transportation Research Part D: Transport and Environment*, vol. 11, no. 4, pp. 292–301, 2006.
- [38] A. Nagurney, Q. Qiang, and L. S. Nagurney, “Environmental impact assessment of transportation networks with degradable links in an era of climate change,” *International Journal of Sustainable Transportation*, vol. 4, no. 3, pp. 154–171, 2010.
- [39] X. Yang, X. J. Ban, and R. Ma, “Mixed equilibria with common constraints on transportation networks,” *Networks and Spatial Economics*, vol. 17, no. 2, pp. 547–579, 2017.
- [40] S. Ryu, A. Chen, and K. Choi, “Solving the stochastic multi-class traffic assignment problem with asymmetric interactions, route overlapping, and vehicle restrictions,” *Journal of Advanced Transportation*, vol. 50, no. 2, pp. 255–270, 2016.
- [41] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Non-linear Equations in Several Variables*, Academic Press, Thailand, 1970.

Research Article

Predicting and Visualizing the Uncertainty Propagations in Traffic Assignments Model Using Monte Carlo Simulation Method

Mundher Seger  and Lajos Kisgyörgy

Highway and Railway Department, Budapest University of Technology and Economics, 1111 Budapest, Hungary

Correspondence should be addressed to Mundher Seger; mundher.seger@gmail.com

Received 20 July 2018; Accepted 10 September 2018; Published 1 October 2018

Guest Editor: Lele Zhang

Copyright © 2018 Mundher Seger and Lajos Kisgyörgy. 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.

Uncertainty can be found at all stages of travel demand model, where the error is passing from one stage to another and propagating over the whole model. Therefore, studying the uncertainty in the last stage is more important because it represents the result of uncertainty in the travel demand model. The objective of this paper is to assist transport modellers in perceiving uncertainty in traffic assignment in the transport network, by building a new methodology to predict the traffic flow and compare predicted values to the real values or values calculated in analytical methods. This methodology was built using Monte Carlo simulation method to quantify uncertainty in traffic flows on a transport network. The values of OD matrix were considered as stochastic variables following a specific probability distribution. And, the results of the simulation process represent the predicted traffic flows in each link on the transport network. Consequently, these predicted results are classified into four cases according to variability and bias. Finally, the results are drawn into figures to visualize the uncertainty in traffic assignments. This methodology was applied to a case study using different scenarios. These scenarios are varying according to inputs parameters used in MC simulation. The simulation results for the scenarios gave different bias for each link separately according to the physical feature of the transport network and original OD matrix, but in general, there is a direct relationship between the input parameter of standard deviation with the bias and variability of the predicted traffic flow for all scenarios.

1. Introduction

Forecasting of travel demand represents the fundamental step of planning and management of transportation facilities [1]. These forecasts are subjected to various sources of errors including error in the measurement of input data, error in the estimated value of model parameters, and error in the specification of the underlying models themselves; also the model itself may be stochastic, and the scenarios adopted for model forecasting may not necessarily be compatible with the real evolution of the transport system [2, 3]. Studying of the uncertainty of travel demand not only is aimed at stripping the values of traffic flow from errors but mainly aims to know the effect of these errors regarding the nature of the variability and bias on traffic flow and the likelihood of occurrences; this will be discussed in this study.

The main goal of travel demand model is traffic forecasting in different stages; generation, distribution, and assignment are to determine future values of the model output variables that are associated with a specific combination of input variables [1, 4]. However, it is impossible to give an exact prediction; no model can be constructed to provide 100% accurate predictions of the future behaviour of a system. A prediction should handle uncertainties by treating output variables stochastically. Without the additional information provided by probability analysis, there is no solid evidence for comparing the predicted value to real value or another prediction [5, 6]. As a result, any method used for prediction should include an assessment of the uncertainty in the predicted values. Uncertainty in traffic forecasts is inherent significantly in travel demand models. Accordingly, a solid understanding of the degree of uncertainty is helpful not only

for the model designation process but also for the sampling process and the value of precise model results for policy scenario development and strategy evaluation [3, 7, 8].

Although many researchers have studied uncertainty in the travel demand models, only a few studies analysed the impact of the error propagation in whole four-stage sequential transport model frameworks. For example, Zhao and Kockelman [9] found that the uncertainty increased throughout the first three model steps and declined in the traffic assignment step. They argued that this reduction might be due to the network congestion effects on the trip assignment equilibrium system, meaning that capacity limitations might reduce the variability of the outcomes at the link traffic flows. However, they also found out that the reduction of uncertainty in the traffic assignment might also be the result of the accumulation on the same links of independent trips related to various OD pairs. De Jong et al. [6, 10] examined the uncertainty of the Dutch national model and used the standard deviations and correlations of 20-year moving averages of some input data to obtain values from a multivariate normal probability distribution function; they found that congestion reduced final model output uncertainty but only to a minor degree. Ziems et al. [11] analysed the effects of the randomness on different traffic characteristics. They observed that the variability of traffic flow is larger during the congested period of the day and there is a greater variability in the individual corridor rather than aggregated subareas. Manzo et al. [12, 13] stated that the congestion in transportation network does not show a high uncertainty effect on the final output of transport model. Hence, the final uncertainty of traffic flow for links with higher volume/capacity ratio showed a lower dispersion around the base uncertainty value. Rasouli and Timmermans [14, 15] investigated uncertainty of OD matrix using the Dutch national transport model and found that higher levels of traffic flow result in lower levels of uncertainty for different model output. Thus, the researcher emphasized that the degree of uncertainty grows higher if the focus of attention shifts from aggregate system performance indicators through OD matrix to disaggregated space-time sequences and performance indicators.

Future year travel demand forecasting is not an exact science, and there are complicated underlying mechanisms that inherently generate uncertainty in the forecasts. Modelling these complicated mechanisms requires numerous variables and behavioural components whose variability may be poorly determined or simply ignored. In this case, it is illogical to take a single view of the future without considering the uncertainty in travel demand modelling. Thus, to provide more efficient and reliable transport solutions for future, transport analysts and planners have to observe and predict uncertainty in transport systems [3, 8, 16].

Uncertainty becomes relevant in transportation modelling not only in case of diverging views such as if risks are very high if the policy is controversial and if there are concerns about model limitations, but also in case of certain views: several points estimated based on different scenarios are given to account for uncertainty [17]. Ideally, analysts would wish to understand the separate and collective impact of these errors on the uncertainty of model forecasts, to be

able to attach credible confidence intervals to model forecasts and optimize the allocation of study resources [8]. However, in large model systems, the interaction between each of these sources of error can be very complicated, making the analysis of propagation of uncertainty through the modelling process extremely challenging [18, 19]. Nevertheless, the increased participation in recent years of the private sector in the delivery of transport infrastructure projects has raised the requirement for accurate traffic demand forecasts and led to renewed interest in the analysis of model uncertainty [20].

This study presents a new methodology for exploring nature of uncertainty propagation deriving from input OD matrix in a four-stage transport model using Monte Carlo simulation method. MC method was used to generate data for OD matrix for three types of probability distributions (i.e., normal, lognormal, and extreme value). Then, the generated OD matrices were entered into the VISUM software to predict traffic assignment attributes. Moreover, the final step was analysing the visualizing the variability and bias of predicted for traffic flow in the transport network links.

The current study contributes to the present literature of uncertainty in transportation planning, primarily by (i) developing a methodology to predict the uncertainty in transport network depending on the variability of input OD matrix, (ii) examining the uncertainty impact on transport model by using different probability distributions in the input data, (iii) adopting a new method to visualize the uncertainty according to a probability of occurrences, and (iv) investigating the probability distributions of output traffic flow on transport network depending on the probability distributions in input data.

2. Methodology

A new methodology was developed for quantifying and characterizing predictive uncertainty in traffic assignment models. The structure of this work directly supports a visual segmentation of uncertainty for transport network to present error and bias in traffic volumes calculated by traffic assignment models. This methodology consists of five stages: (i) input stage; (ii) MC simulation process stage; (iii) analysis of predicted traffic flow stage; (iv) predictive uncertainty stage; (v) uncertainty visualization stage. The relationships connecting these stages of the methodology are presented in Figure 1, and the mathematical and logical computations of this methodology are illustrated in an algorithm (see Algorithm 1).

2.1. Input Stage. The principal task in predictive modelling is to estimate the behaviour of a modelling function, in this case, traffic assignment function $f_{TO \rightarrow D}$ defined to calculate the traffic flow between Oregon zone (O) to Destination zone (D). This work addresses the case where $f_{TO \rightarrow D}$ can be calculated at a finite set of iterations (N_i). Monte Carlo simulation method was used to generate the input data with the parameters: standard deviation (σ) and mean (μ) to produce finite OD matrices. The required data consist of three parts: the first part is setting physical features of the transport network (TN) and traffic analysis zones (TAZ)

Require: $Z_{i=1,n}; \forall Z_i \in \mathbb{N}$	▷ Zone definition.
$O_{i=1,n}; O_i \subset Z$	▷ Origin definition.
$D_{i=1,n}; D_i \subset Z$	▷ Destination definition.
$n; \forall n \in \mathbb{N}$	▷ number of zones.
$OD_{(n,n)}; \forall OD_{(n,n)}$	▷ Matrix of Origin Destination (OD).
$OD_{(i,i)}^{obs}; \forall OD_{(i,i)}^{obs} \in \mathbb{R}$	▷ Observed value of OD pair.
$f_{T \rightarrow D}$	▷ Traffic Assignment function.
$x_{m=1,M}; \forall x_m \in \mathbb{R}$	▷ Observed values of traffic assignment function.
$M; \forall M \in \mathbb{N}$	▷ Number of Links of the transport network.
PD	▷ Define the Probability Distribution type.
$\mu, \sigma; \forall \mu, \sigma \in \mathbb{R}$	▷ Parameters range for the probability distribution.
$N_s; \forall N_s \in \mathbb{N}$	▷ Number of iterations.
(01) $\{\varphi_k\}_{k=1}^{N_s} \leftarrow MC[PD(\mu, \sigma)]$	▷ Get samples by the defined probability distribution and parameters using Monte Carlo procedure.
(02) For $k = 1$ to N_s do	
(03) For $i = 1$ to n do	
(04) For $j = 1$ to n do	
(05) $OD_{(i,i)}^k \leftarrow OD_{(i,i)}^{obs} * \{\varphi_k\}$	▷Get OD matrix randomly.
(06) end for	
(07) end for	
(08) end for	
(09) For $k = 1$ to N_s do	▷Run Visum.
(10) $\chi_m^k \leftarrow f_{T \rightarrow D}(OD_{(i,i)}^k); \forall \chi_m^k \in \mathbb{R}$	▷Get result attribute from traffic assignment function.
(11) end for	
(12) For $m = 1$ to M do	
(13) For $k = 1$ to N_s do	
(14) $GEH_m^k \leftarrow \sqrt{\frac{2(\chi_m^k - x_m)^2}{(\chi_m^k + x_m)}}$	▷Get the value of GEH statistic for each iteration.
(15) end for	
(16) $GEH_m \leftarrow \frac{\sum_{k=1}^{N_s} GHE_m^k}{N_s}; \forall GHE_m \in \mathbb{R}$	▷Get the average value of GEH statistic for each link.
(17) $\bar{\chi}_m \leftarrow \frac{\sum_{k=1}^{N_s} \chi_m^k}{N_s}$	▷Get the average for each link.
(18) $\eta_m \leftarrow (\bar{\chi}_m - x_m); \forall \eta_m \in \mathbb{R}$	▷Get the total bias for each link.
(19) $\sigma_m \leftarrow \sqrt{\frac{1}{N_s} \sum_{k=1}^{N_s} (\chi_m^k - \bar{\chi}_m)^2}; \forall \sigma_m \in \mathbb{R}$	▷Get the standard deviation for each link.
(20) $c_l \leftarrow x_m + 6.25 - \frac{\sqrt{100x_m + 156.25}}{2}$	▷Get the bias lower limit for each link.
(21) $c_u \leftarrow x_m + 6.25 + \frac{\sqrt{100x_m + 156.25}}{2}$	▷Get the bias upper limit for each link.
(22) end for	
(23) For $m = 1$ to M do	
(24) If $GEH_m < 5$	Then end for
(25) For $k = 1$ to N_s do	▷Check the error and the bias for each iteration.
(26) If $\bar{\chi}_m - \sigma_m \leq \chi_m^k \leq \bar{\chi}_m + \sigma_m$	Then low variety
	Else high variety
(27) If $c_l \leq \chi_m^k \leq c_u$	Then low bias
	Else high bias
(28) end for	
(29) end for	
(30) End	

ALGORITHM 1: Prediction of uncertainty in traffic assignment using Monte Carlo simulation.

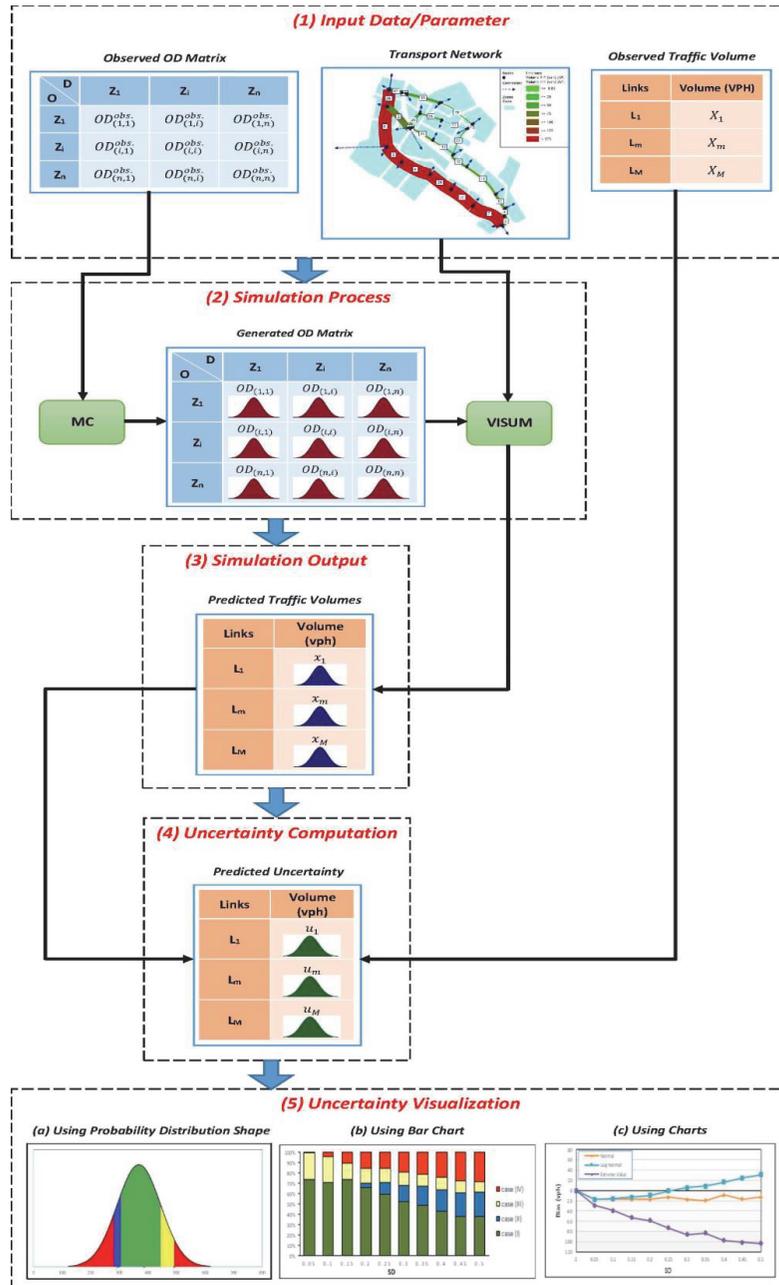


FIGURE 1: The methodology of uncertainty prediction.

by VISUM. The second part is defining the observed OD matrix. Finally, the third part is finding the observed traffic flow attribute either by counting the real value of traffic flow or by running VISUM to get it. Monte Carlo simulation method is used to generate the OD matrices. Therefore, the required parameters for this process are standard deviation (σ), mean (μ), and the number of iterations (N_c). Moreover, the type of probability distribution (PD) must be chosen at the beginning. The required input data and parameters are explained in Algorithm 1.

2.2. Monte Carlo Simulation Stage. Monte Carlo (MC) method or random sampling method is a division of computational mathematics. It is created from the mathematics

concepts for “the frequency approximates the probability.” When the solution for a problem is the occurrence probability of a certain event or an expected value of any variants, a testing method is used to obtain the occurrence frequency of an event or the average value for these variants. MC method is based on the probability model and according to the described process by this model. The results of the simulation test are approximate solutions [21]. MC method plays a fundamental role in characterization and quantification of uncertainty. When the accurate calculation of output uncertainties is needed, then Monte Carlo based analysis is a reliable technique, and it is widely applicable. As a result, its application can be found in virtually all engineering fields. Monte Carlo simulation was usually utilized to observe how

errors or variability of a system can propagate to the final result [22].

In this stage, the simulation code has been written using both Matlab software and Component Object Model (COM) of VISUM software in purpose of predicting traffic flows on the transport network. The simulation code involves two processes: the first is the generation of OD matrix (OD^k) by Monte Carlo method for a given type of probability distribution via observed OD matrix (OD^{obs}).

$$\{\varphi_k\}_{k=1}^{N_s} = MC [PD(\mu, \sigma)] \quad (1)$$

$$OD^k = OD^{obs} * \{\varphi_k\} \quad (2)$$

where $\{\varphi_k\}$ is sample value using Monte Carlo procedure, MC is Monte Carlo procedure, k is iteration number, N_s is number of iterations, PD is probability distribution, μ is mean parameter of PD , σ is standard deviation parameter of PD , OD^k is generated OD matrix, and OD^{obs} is observed OD matrix.

The second process is the process of entering generated OD matrix and the transport network (TN) in the VISUM to produce the traffic flow attribute for the road network from traffic assignment function.

$$\chi_{m=1,M}^k = f_{TO \rightarrow D}(OD^k) \quad (3)$$

where $f_{TO \rightarrow D}$ is traffic assignment function, $\chi_{m=1,M}^k$ is traffic flow attribute, m is link number, and M is number of links.

This simulation is repeated for a finite number of iterations N_s specified by the researcher. With a note, increasing number of iterations N_s leads to improving the accuracy of results but extending simulation time.

2.3. Analysis of Predicted Traffic Flow Stage. One of the most important aspects of a simulation study is an analysis of simulation experiments. In this stage, the results of the simulation process are analysed for both statistical characterization and examining the fitting for the distribution (Goodness-of-Fit Test). The outcomes of the simulation process are traffic flow attributes for all links on the transport network. The number of obtained attributes is equal to the number of iterations N_s used in the simulation.

The statistical characteristics of predictive traffic flow are defined by two parameters: the average value results (μ) and the variability (σ). In this case, the average traffic flow for all links in transport network is shown in (4). Finally, standard deviation represents the variability of results in (5).

$$\bar{\chi}_m = \frac{\sum_{k=1}^{N_s} \chi_m}{N_s} \quad (4)$$

$$\sigma_m = \sqrt{\frac{1}{N_s} \sum (\chi_m - \bar{\chi}_m)^2} \quad (5)$$

where χ_m is traffic flow for the link (m), $\bar{\chi}_m$ is average traffic flow for the link (m), and σ_m is the standard deviation in traffic flow for the link (m).

2.4. Predictive Uncertainty. Meaningful quantification of data and structural uncertainties in conceptual travel demand modelling is a significant challenge in the transportation system. Heterogeneous, insufficient, unstable, and unsteady characterize data used to build OD matrix because of the difficulty in measurements, and the nature of the individual behaviour leads to variation of the parameter of OD matrix in one side. Besides limitations and lack of information in mathematical concepts and the structure in four-stage sequential transport model frameworks lead to variation in traffic flow in transport networks.

The predictive uncertainty is defined by joint consideration of the mean predictive error (i.e., statistical bias) and the predictive variability (i.e., statistical standard deviation). In this methodology, the uncertainty has been predicted by joining the traffic flow attribute from input stage (first stage) and the traffic flow attributes obtained by simulation processes (fourth stage).

The mathematical operations in this stage include both calculating bias in traffic flow as (6) and determining the limitations of allowed biases in the transport network.

$$\eta_m = (\chi_m - x_m) \quad (6)$$

where η_m is bias in traffic volume for the link (m), χ_m is predicted traffic flow for the link (m), and x_m is observed traffic flow for the link (m).

In any uncertainty quantification process, setting limits for the predictive uncertainty is required to increase understanding of the researchers to models behaviour in both bias and predictive variability. The GEH statistic has been used as a limitation of the bias in this study. The GEH statistic is a form of Chi-squared statistic that can be used to compare observed and modelled counts as (7) [1]. It is helpful for these comparisons because it is sophisticated for relative and absolute errors. And, as the estimation of the standard error, the standard deviation statistic was adopted as a limitation for the variability in traffic volumes.

$$GEH = \sqrt{\frac{(M - C)^2}{(M + C)/2}} \quad (7)$$

where M is the modelled traffic flow and C is the observed traffic flow.

GEH statistic bands less than 5 [1] are used to explain bias limit for each link.

$$GEH = \sqrt{\frac{(\chi_m - x_m)^2}{(\chi_m + x_m)/2}} = 5 \quad (8)$$

By solving (8), the upper and the lower bias limit in predicted traffic flow are as follows:

$$c_{l_m} = x_m + 6.25 - \frac{\sqrt{100x_m + 156.25}}{2} \quad (9)$$

$$c_{u_m} = x_m + 6.25 + \frac{\sqrt{100x_m + 156.25}}{2} \quad (10)$$

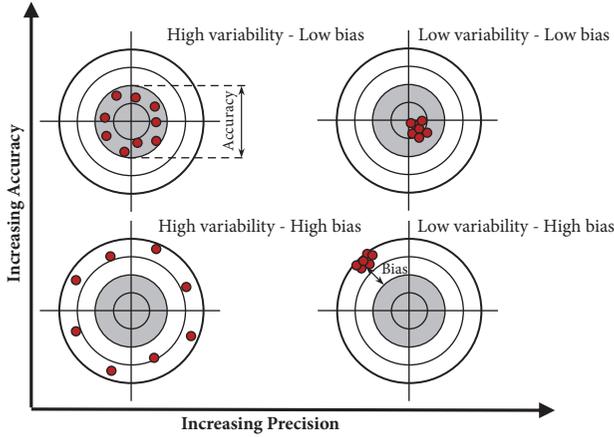


FIGURE 2: The relationship between precision and accuracy.

where χ_m is predicted traffic flow for the link (m), x_m is observed traffic flow for the link (m), c_l is lower accurate limit for the link (m), and c_u is upper accurate limit for the link (m).

The allowed accuracy and variability are presented in (11) and (12), respectively.

$$\text{Accuracy allowable for the link } (m) = c_u - c_l \quad (11)$$

$$\text{Variability allowable for the link } (m) = \mp \sigma \quad (12)$$

If the predicted traffic flow values lie within the standard deviation, then these are precise. Otherwise, these predicted values are imprecise (see (13) and (14)).

$$\begin{aligned} \text{If } \bar{\chi}_m - \sigma_m \leq \chi_m^k \\ \leq \bar{\chi}_m + \sigma_m, \text{ then low variety (precise)} \end{aligned} \quad (13)$$

$$\begin{aligned} \text{If } \bar{\chi}_m - \sigma_m > \chi_m^k \vee \chi_m^k \\ > \bar{\chi}_m \\ + \sigma_m, \text{ then high variety (imprecise)} \end{aligned} \quad (14)$$

In the same way, if the predicted traffic flow values lie within upper and lower bias limits, then these are accurate. Otherwise, these predicted values are biased (see (15) and (16)).

$$\text{If } c_l \leq \chi_m^k \leq c_u, \text{ then low bias (accurate)} \quad (15)$$

$$\text{If } c_l > \chi_m^k \vee \chi_m^k \leq c_u, \text{ then high bias (inaccurate)} \quad (16)$$

To interpret the concept of precision (i.e., the variability of values) and the accuracy (i.e., the bias of values), see Figure 2.

2.5. Visualizing Predictive Uncertainty. The last stage of this methodology is uncertainty visualization. Uncertainty visualization is endeavouring to display data together with additional uncertainty information. These visualizations present a more complete and accurate interpretation of data for

researchers to analyse [23]. Thus, visualization is a useful method for addressing many forms of information uncertainty, and it is a helpful approach to the investigation and communicating of large data sets [24].

Applications that use visual graphs and comparative figures to indicate information variability or draw levels of confidence in data values help analysts better understand and cope with uncertain information better than using digital tables and metadata [25]. Consequently, visualizing the uncertainty is essential for risk analysis and decision-making tasks. However, it is still a challenge, because describing the uncertainty is a complex concept with many interactions, definitions, and interpretations in transportation models. Uncertainty can be introduced into information visualizations as the data is collected, transformed, and integrated into information [25, 26]. In the absence of combined presentation of data and its associated uncertainty, the analysis of the information visualization is incomplete at best and may lead to inaccurate or incorrect conclusions. Therefore, there is a need to display information together with their uncertainty for accurate interpretation and precise decision-making [26, 27].

There are different methods used to visualize uncertainty: statistical and probability-based visualization, point and global visualization, used colours, financial visualization, icons, ontology, lexicon, etc. [20]. In this methodology, statistical and probability-based visualization method is used to visualize uncertainty. This method is one of the most powerful methods to address conceptual model uncertainty with a bar chart, probability distributions, and traditional charts represented by random variables. This method demonstrates the central tendency, dispersion, skewness, and modal characteristics of a random variable. In this methodology, two steps are used to visualize the uncertainty of the predicted traffic flow.

The first step classifies the predicted traffic flow values into four cases according to bias and variability.

Case I (accurate and precise (low predictive uncertainty)). This case occurred when the predicted traffic flows (χ_m) are close to the mean prediction ($\bar{\chi}_m$), i.e., within the standard deviation ($\pm\sigma$). And predicted traffic flows are within the accuracy limits (c_u and c_l). This means that the results are of low variability and low bias.

Case II (accurate and imprecise (moderate predictive uncertainty)). This case occurred when the predicted traffic flows (χ_m) are close to the mean prediction ($\bar{\chi}_m$), i.e., within the standard deviation ($\pm\sigma$). And predicted traffic flows are out of the accuracy limits (c_u and c_l). This means that the results are of low variability and high bias.

Case III (inaccurate and precise (high predictive uncertainty with ensemble agreement)). This case occurred when the predicted traffic flows (χ_m) are far from the mean prediction ($\bar{\chi}_m$), i.e., out the standard deviation ($\pm\sigma$). And predicted traffic flows are within the accuracy limits (c_u and c_l). This means that the results are of high variability and low bias.

TABLE 1: This table shows the classification of the four cases of uncertainties.

		Variability limitations (Precision)	
		Low	High
Bias limitations (Accuracy)	Low	Case I Low variability - Low bias (Low predictive uncertainty)	Case III Low variability - High bias (High predictive uncertainty with ensemble estimates)
	High	Case II Low variability - High bias (Moderate predictive uncertainty)	Case IV High variability - High bias (High predictive uncertainty with divergent estimates)

Case IV (inaccurate and imprecise (high predictive uncertainty with divergent estimates)). This case occurred when the predicted traffic flows (χ_m) are far from the mean prediction ($\bar{\chi}_m$), i.e., out the standard deviation ($\pm\sigma$). And predicted traffic flows are out of the accuracy limits (c_u and c_l). This means that the results are of high variability and high bias.

The second step of the visualization process is giving specific colour for each case of uncertainty. Table 1 illustrates the colours, characterizations, and limitations of the four uncertainty cases. Figure 3 demonstrates the probability density curves of predictive uncertainty cases. In practice, all the four cases can appear with different likelihood proportions. Figure 4 shows an example of a probability density curve for predicted traffic flow for a link (m). The coloured areas under the curve represent the predictive uncertainty cases.

The major role of this visualization of uncertainty is to give information about the level of uncertainty to the decision-makers. Based on these cases they can see how reliable the predictions of the model are and if they can accept the risks relating to the given predictive uncertainty. If predictions that are more precise are needed, then sensitivity analysis helps to identify the input data most dominantly influencing the predictive uncertainty. Then these input data should be measured more precisely.

3. Case Study

The developed methodology has been applied in a small city, Ajka, located in Hungary. The results of case study were presented earlier in [27] for demonstrating the effects of variability of input variables on the model results. Here we used the case study and its results to describe the implementation of the methodology. Figure 5 presents the transport network (NT) and traffic analysis zones (TAZ) for the study area. This case study applied the developed methodology step by step, as follows.

3.1. Input Data and Parameters. The required data include information about the physical feature of the study area, where the number of TAZ=25 and the number of links (M) = 50. Also collecting 625 OD pairs represents the travel distribution between zones (i.e., OD^{obs}). Besides, in this case,

the travel assignment attribute has been gotten by executing the VISUM (i.e., $x_{m=1,M}$).

In this case, three types of probability distributions (PD) have been used: (1) normal distribution, (2) lognormal distribution, and (3) extreme value distribution. Ten values of standard deviations (σ) have been examined for each distribution; these values are ranged from 0.05 to 0.50, while the observed values of the OD matrix have been used as mean values (μ).

3.2. Monte Carlo Simulation. In this simulation, 30 scenarios have been experimented. These scenarios have been grouped into three groups according to probability distribution (PD) type. And for each group, the parameter of standard deviation (σ) has been changed for each scenario. Besides, 1000 iterations have been used per each scenario. Thus, the total number of iterations used in the whole simulation process is as follows: (30 scenarios) x (1000 iterations for each scenario) equal to (30000 iterations). Therefore, this simulation process is expensive; the time spent to complete this simulation is around 300 hours, using a computer with the following specifications: CPU from Intel 8th generation, Core i7 8700K, 6 Cores, 3.70GHz, RAM 32GB, DDR 4. The simulation time of this methodology depends on (1) the number of links on the TN, (2) the number of TAZ, and (3) the number of iterations N_s .

3.3. Analysis of Predicted Traffic Flow. The results of the simulation process are represented by attributes of traffic flow for the links of the transport network. The number of these attributes is equal to the number of simulation iterations. Consequently, the outcomes for each link have been studied separately, by finding the statistical parameters for predictive traffic flows: both of the average value ($\bar{\chi}_m$) and the variability (σ_m). Because there are 50 links on the transport network in this study case, only one link will be presented (we chose link No. 6). For example, Figures 6, 7, and 8 represent three scenarios having a different probability distribution and different standard deviations for the link No. 6.

3.4. Predictive Uncertainty. The obtained data from the previous stage has been processed by Excel-sheet to find the uncertainty in traffic flows. The mathematical calculations

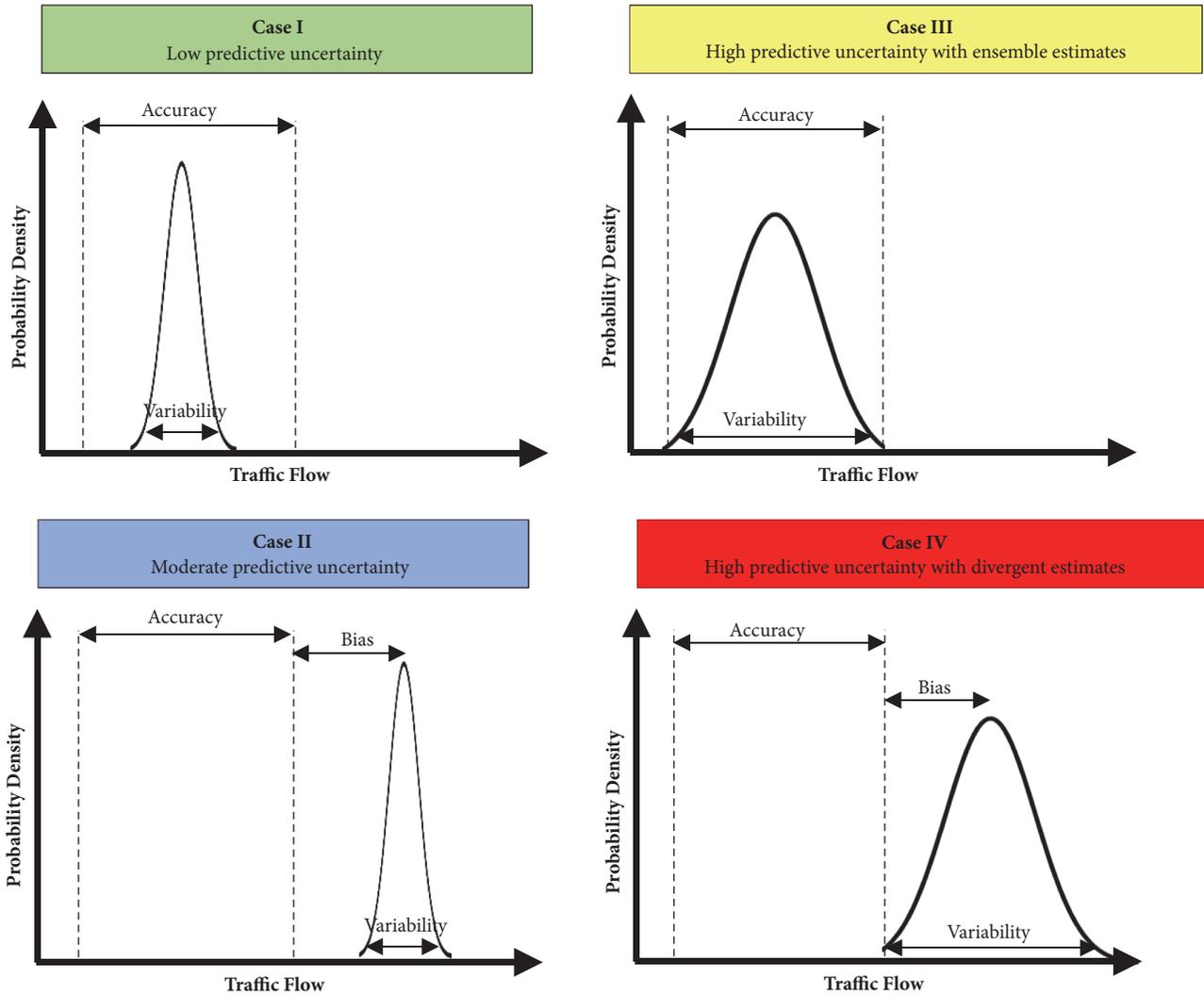


FIGURE 3: The probability density curves of the predicted traffic flow for the four cases.

determine the bias and variability for the predicted traffic flows, as well as the accurate limits for each link separately. For the link No. 6, the lower accurate limit (c_l) = 295 vph and the upper accurate limit (c_u) = 492 vph. Figure 6 shows the percentage of bias in the predicted traffic flows associated with input standard deviations. It can be observed that increasing the standard deviation of inputs leads to increasing bias for the predicted traffic flows, for normal distribution scenarios and lognormal distribution scenarios; there are no or few biases for standard deviations less than 0.10, and the biases increased to reach about 42% in standard deviation equal to 0.50, while for the extreme value distribution scenarios the biases increased in all standard deviations to reach about 54% in standard deviation equal to 0.50. Figure 7 shows the percentage of converging degree for predicted traffic flows joined with input standard deviations. It can be noted that the variability of the predicted traffic flows for normal distribution scenarios ranged between 67.1% and 70.3% for standard deviations less than or equal to 0.40 and decreased after that to reach 57.6% in the standard deviation

equal to 0.50. And, the variability of the predicted traffic flows for lognormal distribution scenarios ranged between 67.9% and 70.3% for standard deviations less than or equal to 0.40 and increased after that to reach 76.4% in the standard deviation equal to 0.50, while the variability of the predicted traffic flows for extreme value distribution scenarios ranged between 67.2% and 72.0% for standard deviations less than or equal to 0.35 and decreased after that to reach 61.4% in the standard deviation equal to 0.50.

3.5. *Visualizing Predictive Uncertainty.* Finally, the probability of the uncertainty cases for the predicted traffic flows in transport network has been visualized into bar-charts by merging Figures 6 and 7. This visualization provides transport planners and engineers with the possibility to monitor and identify which of the links suffers from bias and unexpected change in traffic volumes in the event of a change in the conditions of traffic parameters and experiment different scenarios on the transport network.

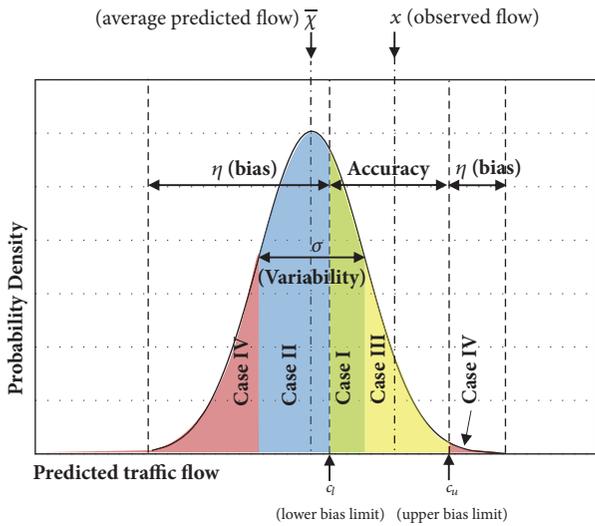


FIGURE 4: The probability density of predicted traffic flows addressed with the uncertainty cases.

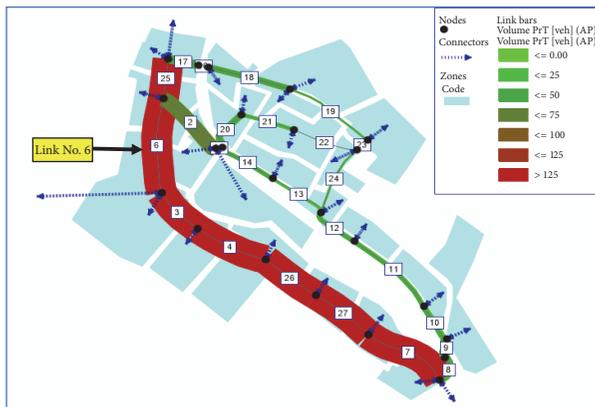


FIGURE 5: Study area (Ajka, Hungary), showing observed traffic flows on the transport network [27].

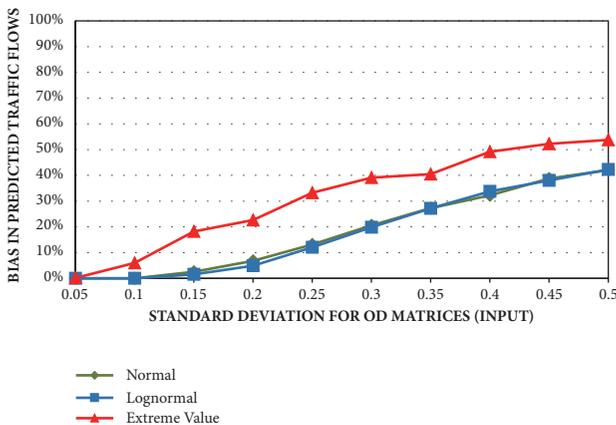


FIGURE 6: The relationship between the standard deviation for OD matrices (input) and the bias in the predicted traffic flows for link No. 6.

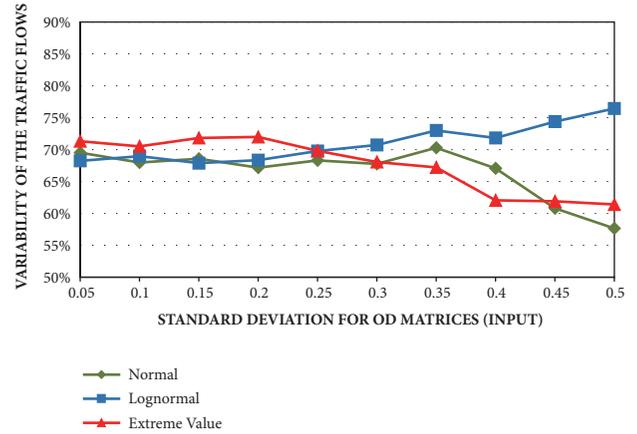


FIGURE 7: The relationship between the standard deviation for OD matrices (input) and the variability of the predicted traffic flows (output) for link No. 6.

The coloured bar-charts give a whole idea about the probability ratios of the uncertainty cases. In general, the percentage of green colour (Case I) means the probability of predicted traffic flow will lie within the allowed limits of accuracy and variability. And the percentage of yellow colour (Case III) means the probability of predicted traffic flow will lie still within the allowed accuracy but outside the allowed variability. Both percentages of green colour (Case I) and yellow colour (Case III) are located within the accuracy range, while the percentage of blue colour (Case II) means the probability of predicted traffic flow will lie outside the allowed accuracy but still within the allowed variability. And, the percentage of red colour (Case IV) means the probability of predicted traffic flow will lie outside the allowed limits of accuracy and variability. Both percentages of blue colour (Case II) and red colour (Case IV) are located outside the accuracy range.

Although the purpose of the research is visualization of the uncertainty, it is good to give more details by text. In the same example, for the link No. 6, concerning normal distribution scenarios as shown in Figure 8, we can see that Case I is decreased from 69.5% in SD=0.05 to 54.7% in SD=0.50; Case II started appearing from SD=0.30 and increased to 5.3% in SD=0.35 and decreased after that to reach 2.9% in SD=0.50; Case III is decreased from 32% in SD=0.05 to 3.2% in SD=0.50; and Case IV started appearing from SD=0.15 and increased to reach 39.1% in SD=0.50.

Similarly, we can interpret Figure 9 for lognormal scenarios; we can see that Case I is decreased from 68.2% in SD=0.05 to 57.7% in SD=0.50; Case II started appearing from SD=0.35 and increased to 18.7% in SD=0.50; Case III is decreased from 31.8% in SD=0.05 to 0.6% in SD=0.35 and disappeared after that; and Case IV started appearing from SD=0.15 and increased to reach 23.6% in SD=0.50.

Likewise, we can interpret Figure 10 for extreme value scenarios; we can see that Case I is decreased from 71.3% in SD=0.05 to 33.3% in SD=0.50; Case II started appearing from SD=0.15 and increased to 28.1% in SD=0.50; Case III is decreased from 28.6% in SD=0.05 to 12.6% in SD=0.50; and

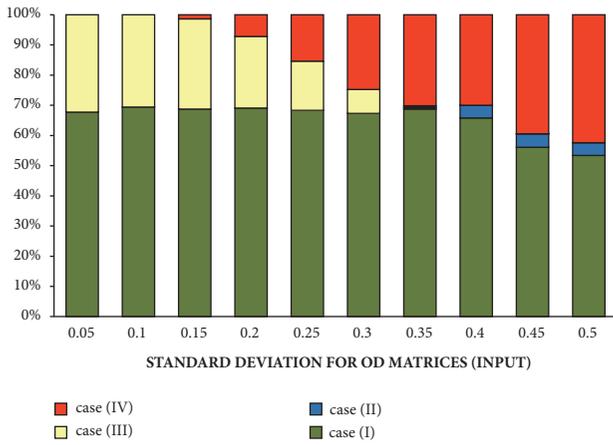


FIGURE 8: The predicted uncertainty for the link No. 6, applying normal distribution scenarios [27].

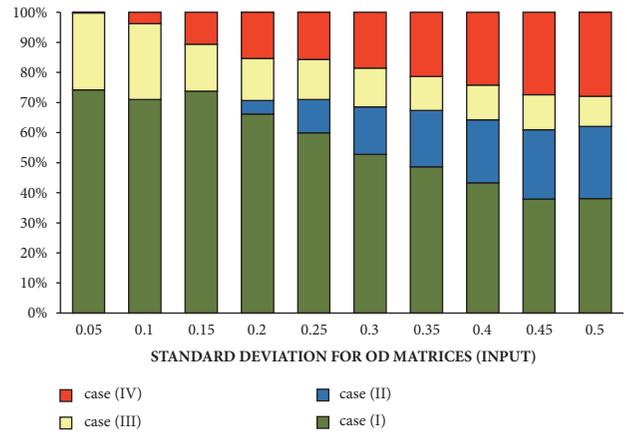


FIGURE 10: The predicted uncertainty for the link No. 6, applying extreme value distribution scenarios [27].

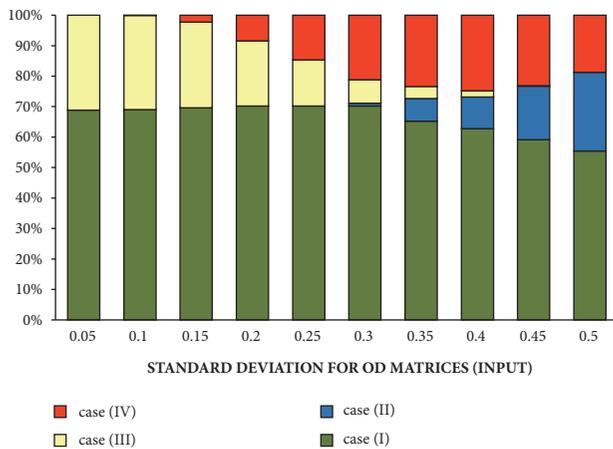


FIGURE 9: The predicted uncertainty for the link No. 6, applying lognormal distribution scenarios [27].

Case IV started increasing from 0.6% in SD=0.05 to 25.7% in SD=0.50.

Figures 8, 9, and 10 show that the predictive uncertainty strongly depends on the uncertainties of the input data. This makes the application of stochastic approaches in the practice of transport modelling highly advisable. Quantifying the uncertainty levels of predictions indicates the acceptability of the predictions and helps to identify the relating risks.

4. Conclusions

Visualization of predictive uncertainties helps to understand the stochastic nature of predictions. To be able to make decisions based on model predictions, decision-makers should have information about the accuracy and precision of the predictions. If the predictive uncertainty of the model is not acceptable, then sensitivity analysis helps to identify the input data most dominantly influencing the predictive uncertainty.

In this paper, a new methodology has been presented to predict traffic flow and visualize the uncertainty in those

predicted values. This methodology enables applying various scenarios showing the variation in traffic flow on transport network by supposing that the input values of OD matrix are varying according to a specific probability distribution. The importance of this methodology is permitting transport planners and decision-makers to monitor and identify which of the links suffers from bias and unexpected change in traffic volumes in the event of a change in the conditions of inputs OD matrix.

The algorithm of this methodology consists of two parts: the first part has been built on Monte Carlo simulation method to generate numerous OD matrices, and VISUM software for getting the traffic assignment on a transport network. The results of this part represent predicted traffic flows on each link of the transport network. These predicted traffic flows suffer from uncertainty in both a bias from the observed value and variability from the average predicted value, while the second part of the algorithm was designed to categorize the uncertainty of the predicted traffic flows into four cases according to variability and bias: Case I (low variability, low bias), Case II (low variability, high bias), Case III (high variability, low bias), and Case IV (high variability, high bias). Finally, the percentages of these cases have been visualized in coloured bar-charts. The percentage of each case represents the likelihood of occurring (i.e., the likelihood of the predicted traffic flow to biasing or varying depends on the percentages of these cases).

Finally, the methodology has been tested in a small study area using three main-scenarios; each of them has (i) different probability distribution (normal, lognormal, and extreme value) and (ii) 10 subscenarios different according to standard deviation parameter graded from 0.05 to 0.50. The obtained results of this study area showed that uncertainty in traffic flow is found on all links of the transport network but in different degrees, depending on the scenario's parameters and the observed traffic flow.

The current case study shows that the effect of applying scenarios had the same simulation parameters for all zones. Future research will consider applying different simulation

parameters in the same scenario according to land use characteristic of each zone and how the accuracy and precision of the predicted traffic flows can be improved once the case of uncertainty of the predicted traffic flow is known.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request. The software used in this study can be requested from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

- [1] J. d. Ortúzar and L. G. Willumsen, *Modelling Transport*, John Wiley and Sons Ltd Publication, Chichester, UK, 4th edition, 2011.
- [2] H. S. Mahmassani, "Uncertainty in transportation systems evaluation: Issues and approaches," *Transportation Planning and Technology*, vol. 9, no. 1, pp. 1–12, 1984.
- [3] M. Ottomanelli and C. K. Wong, "Modelling uncertainty in traffic and transportation systems," *Transportmetrica*, vol. 7, no. 1, pp. 1–3, 2011.
- [4] Cambridge Systematics Inc., Vanasse Hangen Brustlin Inc., Gallop Corporation, C. R. Bhat, Shapiro Transportation Consulting LLC, and Martin Alexiou Bryson PLLC, *Travel Demand Forecasting: Parameters and Techniques*, 2012.
- [5] W. Briggs, *Uncertainty: The Soul of Modeling, Probability and Statistics*, Springer International Publishing, Switzerland, 2016.
- [6] G. de Jong, A. Daly, M. Pieters, S. Miller, R. Plasmeijer, and F. Hofman, "Uncertainty in traffic forecasts: Literature review and new results for The Netherlands," *Transportation*, vol. 34, no. 4, pp. 375–395, 2007.
- [7] G. J. De Jong, A. Daly, M. Pieters, S. Miller, and F. Hofman, "Uncertainty in traffic forecasts: literature review and new results for the Netherlands," *European Journal of Transport and Infrastructure Research (EJTIR)*, pp. 1–23, 2005.
- [8] A. Matas, J.-L. Raymond, and A. Ruiz, "Traffic forecasts under uncertainty and capacity constraints," *Transportation*, vol. 39, no. 1, pp. 1–17, 2012.
- [9] Y. Zhao and K. M. Kockelman, "The propagation of uncertainty through travel demand models: an exploratory analysis," *Annals of Regional Science*, vol. 36, no. 1, pp. 145–163, 2002.
- [10] D. J. Van Der Laan, E. De Jonge, and J. Solcer, "Effect of displaying uncertainty in line and bar charts: Presentation and interpretation," in *Proceedings of the 6th International Conference on Information Visualization Theory and Applications, IVAPP 2015*, pp. 225–232, Germany, March 2015.
- [11] S. E. Ziems, B. Sana, J. Plotz, and R. M. Pendyala, "Stochastic variability in microsimulation modeling results and convergence of corridor-level characteristics," in *Proceedings of the Transportation Research Board 90th Annual Meeting*, vol. 11-3560, 2010.
- [12] S. Manzo, O. A. Nielsen, and C. G. Prato, "Investigating uncertainty in transport modelling: a four-stage model case study," *Latsis Symposium*, 2012.
- [13] S. Manzo, O. A. Nielsen, and C. G. Prato, "How uncertainty in input and parameters influences transport model: Output A four-stage model case-study," *Transport Policy*, vol. 38, pp. 64–72, 2015.
- [14] S. Rasouli and H. Timmermans, "Uncertainty in travel demand forecasting models: Literature review and research agenda," *Transportation Letters*, vol. 4, no. 1, pp. 55–73, 2012.
- [15] S. Rasouli and H. Timmermans, "Applications of theories and models of choice and decision-making under conditions of uncertainty in travel behavior research," *Travel Behaviour and Society*, vol. 1, no. 3, pp. 79–90, 2014.
- [16] A. Pradhan and K. M. Kockelman, "Uncertainty propagation in an integrated land use-transportation modeling framework: Output variation via UrbanSim," *Transportation Research Record*, no. 1805, pp. 128–135, 2002.
- [17] W. J. Rider, J. R. Kamm, and V. G. Weirs, "Verification, validation and uncertainty quantification workflow in CASL," *Sandia National Laboratories*, no. SAND2010234P, pp. 1–27, 2010.
- [18] W. Walker, P. Harremoës, J. Rotmans et al., "Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support," *Integrated Assessment*, vol. 4, no. 1, pp. 5–17, 2003.
- [19] S. I. Gass and M. C. Fu, *Encyclopedia of Operations Research and Management Science*, Springer US, Boston, MA, USA, 2013.
- [20] B. M. Ayyub and G. J. Klir, *Uncertainty Modeling and Analysis in Engineering and the Sciences*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2006.
- [21] Y. Du, Y. Geng, and L. Sun, "Simulation model based on monte carlo method for traffic assignment in local area road network," *Frontiers of Architecture and Civil Engineering in China*, vol. 3, no. 2, pp. 195–203, 2009.
- [22] H. Janssen, "Monte-Carlo based uncertainty analysis: sampling efficiency and sampling convergence," *Reliability Engineering & System Safety*, vol. 109, pp. 123–132, 2013.
- [23] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha, "Approaches to uncertainty visualization," *The Visual Computer*, vol. 13, no. 8, pp. 370–390, 1997.
- [24] H. Griethe and H. Schumann, "The visualization of uncertain data: methods and problems," in *Proceedings of the SimVis '06*, vol. vi, pp. 143–156, SCS Publishing House, 2006.
- [25] L. Gosink, K. Bensema, T. Pulsipher et al., "Characterizing and visualizing predictive uncertainty in numerical ensembles through bayesian model averaging," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2703–2712, 2013.
- [26] J. Chang-Sung and A. Pang, "Reconfigurable disc trees for visualizing large hierarchical information space," in *Proceedings of the IEEE Symposium on Information Visualization*, pp. 19–25, IEEE, Research Triangle, Calif, USA, 1998.
- [27] M. Seger and L. Kisgyorgy, "Visualize the effect of input variability on model output in traffic assignment," in *Proceedings of the Fifth International Conference on Road and Rail Infrastructure*, pp. 1233–1240, 2018.

Research Article

Enhancing Freeway Safety through Intervening in Traffic Flow Dynamics Based on Variable Speed Limit Control

Jinming You,¹ Shouen Fang,¹ Lanfang Zhang,¹ John Taplin,² and Jingqiu Guo ¹

¹Key Laboratory of Road and Traffic Engineering of The Ministry of Education, Tongji University, Shanghai 201804, China

²Business School, The University of Western Australia, WA, 6009, Australia

Correspondence should be addressed to Jingqiu Guo; guojingqiu@hotmail.com

Received 19 July 2018; Accepted 9 September 2018; Published 26 September 2018

Guest Editor: Zhiyuan Liu

Copyright © 2018 Jinming You 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.

New technologies and traffic data sources provide great potential to extend advanced strategies in freeway safety research. The High Definition Monitoring System (HDMS) data contribute comprehensive and precise individual vehicle information. This paper proposes an innovative Variable Speed Limit (VSL) based approach to manage crash risks by intervening in traffic flow dynamics on freeways using HDMS data. We first conducted an empirical analysis on real-time crash risk estimation using a binary logistic regression model. Then, intensive microscopic simulations based on AIMSUN were carried out to explore the effects of various intervention strategies with respect to a 3-lane freeway stretch in China. Different speed limits with distinct compliance rates under specified traffic conditions have been simulated. By taking into account the trade-off between safety benefits and delay in travel time, the speed limit strategies were optimized under various traffic conditions and the model with gradient feedback produces more satisfactory performance in controlling real-time crash risks. Last, the results were integrated into lane management strategies. This research can provide new ideas and methods to reveal the freeway crash risk evolution and active traffic management.

1. Introduction

There is a growing body of evidence confirming a positive relationship between the road safety benefits and vehicle speed enforcement, especially on freeways. In China, for example, Shanghai and Jiangsu with intensive freeway networks are actively employing intelligent technology systems for coordinating traffic flow and improving road safety. Previous studies have highlighted the higher vehicle speed on freeways associated with increased crash risk and injury severity [1, 2]. Meanwhile, speed variation among vehicles can disturb traffic flow and create more conflict situations [3]. Active traffic management (ATM) has been emerging in recent years aiming to provide traffic control to improve traffic flow and reduce congestion on freeways. Proper traffic control can significantly reduce delays and improve traffic distribution at a bottleneck, especially under congested [4] and work zone conditions [5, 6]. As a key application of ATM, Variable Speed Limit (VSL) systems aim to dynamically regulate freeway speeds based on real-time traffic flow information.

In the last decades, VSL has been intensively investigated on two main directions: traffic enhancement and safety improvement [7]. For instance, Hegyi et al. [8] proposed the macroscopic traffic flow model METANET with coordinated control of ramp metering and VSL to minimize the total time vehicles spent on the road; the method significantly reduced congestion. Naïve and Empirical Bayes are used to evaluate the effects of the VSL system and results indicate that VSL reduces crashes by 4.5% to 8% [9]. However, instead of mandatory, the advisory VSL does not show significant impact on traffic flow [10]. Especially under low speed limits, some drivers tend to violate the limits in pursuit of their personal benefits. Hence low speed limits may widen the range of flows under homogeneous traffic and contribute to a raise in lane changes [11]. Recent studies have shown that lane-changing maneuvers are a major source of traffic disturbance on a multilane freeway. Therefore, a proper setting is essential in VSL strategies. Instead of just using VSL before or during periods of high congestion, it can be applicable during off-peak periods as well [12]. So far, limited

studies have been focusing on enhancing freeway safety by intervening in traffic flow dynamics based on VSL.

This study aims to apply real-time crash prediction in traffic control management. Previous studies on crash precursors have employed different kinds of traffic data such as loop detectors [1, 2, 13–15], automatic vehicle identification [16, 17], traffic counter data [18], and weather [19, 20] as well as road geometry data [21]. Different data mining and detection methods have been utilized to fully investigate the interrelationship between crash risk and traffic operation data. Abdel-Aty et al. [1] developed a neural network-based classifier to evaluate rear-end crash risk with traffic parameters from five stations: it can identify 75% of the crashes, with false positives of 34%. Ahmed et al. [16] exploited automatic vehicle identification (AVI) data and the model achieves an accuracy of 75.93% and 72.92% for rear-end and all crashes, respectively. As large numbers of false positives might affect drivers' compliance with the system and reduce its effectiveness, various refining approaches have been employed to optimize the evaluation algorithm. Traffic data with high resolution and multiple sources are needed for a better evaluation model with higher predictive accuracy and robustness. Ahmed et al. [17] enhance the AVI data with remote traffic microwave sensor data and their model can successfully identify 88.9% of the crashes with a false positive rate of only 6.5%. Kwak et al. [22] found that traffic flow characteristics leading to crashes differ by segment type and traffic flow state.

However, with respect to data type and resolution, the detectors are often limited and traffic data from continuous detectors cannot be collected or the collected data do not meet the requirements of the models. For instance, in China, detectors are installed far apart on freeways and most segments have not been equipped with detectors or surveillance devices. Regarding to method, generalized linear models could provide direct evidence of the traffic parameters' impacts on crash risk. When dealing with highly nonlinear relationships between traffic flow and crash, it requires more computational, flexible, and nonlinear models [23]. Recent studies found that nonlinear models are capable of achieving higher crash prediction accuracy with less false positives. However, the limitations of the available nonlinear models include heavy computations to reveal deeper connections between the traffic parameters as well as for model calibration. Meanwhile, few studies have thoroughly investigated the application of real-time crash risk prediction. Based on an accurately quantified crash risk evaluation, proper traffic management strategies can be applied and therefore improve road safety. For example, Yu et al. [7] proposed a VSL control algorithm in mountainous freeways and the result indicated a positive outcome in crash risk control. However, most VSL control studies are not safety-oriented and so the only parameter utilized in the crash risk model is speed. Meanwhile, generally it is hard to obtain accurate speed variation data, and the effect of speed dispersion on traffic safety has not been intensively investigated [3].

New technologies and traffic data sources provide great potential to extend advanced strategies in freeway safety research. For example, the High Definition Monitoring System (HDMS) data contribute comprehensive and precise

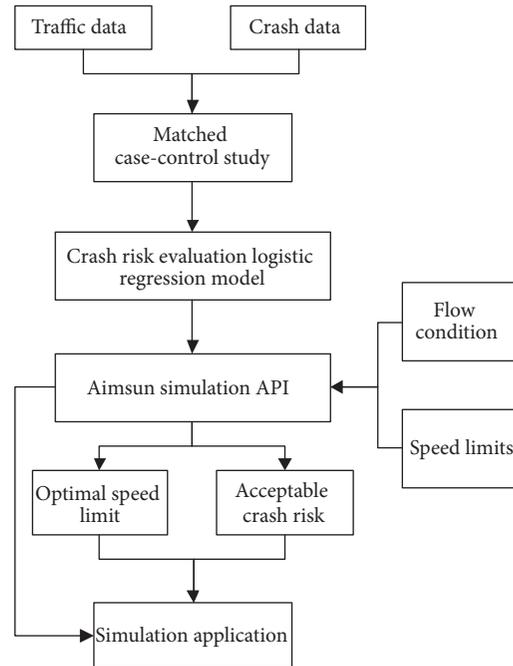


FIGURE 1: Architecture of this study.

individual vehicle information, including vehicle type, speed, lane number, and plate number, as well as high quality photos captured by advanced vehicle license plate recognition systems. In China, HDMS have been installed on major freeways for public security management. The major contributions of this paper consist of the following aspects (Figure 1): (1) to employ HDMS data with individual vehicle information to study the crash mechanism on freeways; (2) to develop a Logistic Regression model for real-time crash risk estimation; (3) to evaluate safety benefits of the optimized VSL based on enhanced AIMSUN simulations on a 100 km freeway stretch; (4) to investigate sensitivities of VSL impacts on driver compliance.

2. Crash Risk Model

2.1. Data Preparation. The study area is G15 Freeway in Nantong, Jiangsu Province, China, with a total length of approximately 100 km, from Sutong Bridge to Fuan Toll (as shown in Figure 2).

Data are obtained from the Public Security Traffic Managing System of the Traffic Management Research Institute, Ministry of Public Security. The freeway is a 6-lane one (3 lanes in each direction). The primary dataset includes all crash data and HDMS data from January to October 2016. The extracted HDMS data cover the lane number, direction, vehicle type and speed, recorded time of vehicle passing, etc. The study area includes five pairs of HD cameras.

The raw crash dataset includes 5924 crashes. However, the majority of crashes are not recorded with detailed location or direction information. 88% of the crashes are involved with multivehicles. Among them, 96% of the crashes are recorded with causes of hitting the fixed objects such as the guardrails

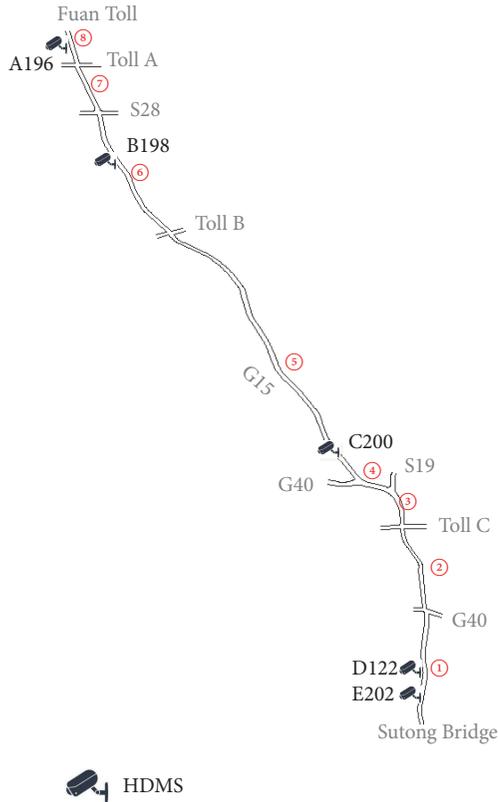


FIGURE 2: The study area and subsegment division.

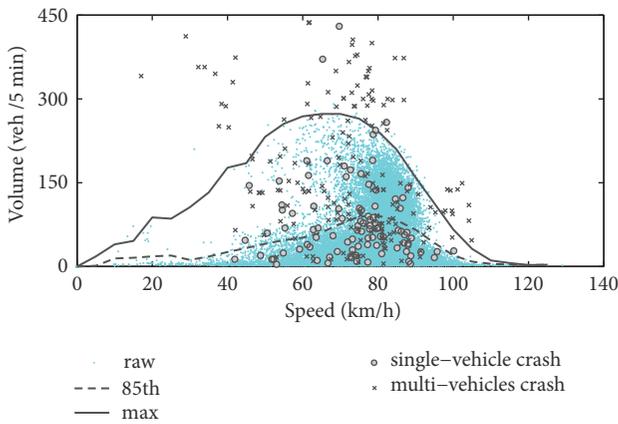


FIGURE 3: The traffic status 5-10 minutes prior to crashes.

and the medians, or hitting the unfixed objects such as the crash barriers. In order to investigate the impact of traffic dynamics on crashes, the traffic status prior to crashes has been examined as Figure 3. In Figure 3, the raw HDMS data are aggregated with the interval of 5 minutes. The datasets from April 1st to May 31st in 2016 have been utilized to show the relationship between traffic volume and speed. The solid curve “max” shows the margin of the data during this period and the dash curve shows the 85th margin of the data during this period. The dash curve demonstrates that, under most conditions, the speed/volume distribution is within the curve. The speed/volume data 5-10 minutes prior to each crash have

TABLE 1: Summary statistics of variables.

Variable	Mean	S.D.	First Quartile	Third Quartile
Q (veh)	103.59	84.80	43.00	138.00
V (km/h)	80.37	12.83	75.04	87.93
DV (km/h)	22.28	5.40	18.73	26.08

been plotted in Figure 3 as well. The symbol “o” refers to the crashes involved with single vehicle and the symbol “x” refers to the crashes involved with multivehicles.

First, the “max” curve shows the similar trend with that of the traditional capacity/speed curve. However, as under most conditions, the traffic state is normal and it is difficult to obtain the saturated flow state with different speeds. Hence, the “max” curve mainly reflects the nonfree flow state, in which the volume is approaching the maximum capacity. The area beyond the “max” curve reflects the chaos flow state or congestion state. Second, comparing the state of single-vehicle crashes and the state of multivehicles we could find that the single-vehicle crashes are more likely to occur within the 85th curve; i.e., the crashes are likely to occur under free flow conditions. This is consistent with several existing studies [24]. Another finding is that the number of multivehicle crashes is in majority in the area beyond the “max” curve. It indicates that when the traffic state approaches congestion state, the crashes are likely to involve multivehicles; especially when the speed is below 60 km/h, only multivehicle crashes are detected. Meanwhile, under all conditions, when the speed is below 40km/h, no single crashes have been recorded.

Additionally, as single-vehicle crashes are usually caused by random effects, such as driving distraction and breaking down, only multivehicles crashes with detailed temporal and spatial information are utilized in this study to investigate the relationship within traffic dynamics and crash risk. The data 5-10 minutes prior to the crashes are utilized to represent the traffic status prior to crashes. The method is commonly used in existing studies [15] as well.

A matched case-control method is utilized to extract the related samples for each sampled crash. A 4:1 control-case ratio is used, as recommended in several previous studies [2]. Each crash case and noncrash case is matched with corresponding traffic data on the same road segment. The four control samples are selected from 14 days before the recorded crash time, 7 days before, 7 days after, and 14 days after. Considering the transferability of the model, the HDMS data 5-10 minutes prior to the samples are aggregated as traffic flow, mean speed, and speed dispersion, which are labeled as Q, V, and DV, respectively.

Some filtering rules are also applied to select the available samples. Due to occasional HDMS system failure, some samples would be matched with invalid HDMS data or missing HDMS data. Noise and outliers are removed from the final dataset. Finally the crash dataset contains 171 samples and the control dataset has 618 non-crash samples. The summary statistics of variables are listed in Table 1.

2.2. Binary Logistic Regression Model. Logistic regression analysis is commonly used to quantify the crash risk in

TABLE 2: Crash risk evaluation model for the whole segment.

Variables	<i>B.</i>	<i>S.E.</i>	<i>Wals.</i>	<i>Df.</i>	<i>Sig.</i>
<i>Q</i>	0.012	0.001	101.806	1	0.000
<i>DV</i>	0.120	0.020	37.882	1	0.000
<i>Constant</i>	-5.487	0.527	108.473	1	0.000
<i>AUC</i>	0.755	<i>Nagelkerke R²</i>		0.288	

TABLE 3: Comparison of the performance of the three models.

Data type	Detector density	Modeling technique	Prediction accuracy
AVI	3 detectors per 2.35 km	Bayesian	0.759
Loop detector	6 detectors per 5.46 km	Genetic Programming	0.608
HDMS	Single detector per 20 km	Logistic Regression	0.755

real-time crash analysis. The traffic condition can be divided into two parts, crash cases ($y_i = 1$) and noncrash cases ($y_i = 0$) with respective probability p_i and $1 - p_i$. The probability of a crash occurrence is estimated by

$$P(y_i = 1 | x_i) = p_i = \frac{e^{\beta_0 + \sum \beta_k x_{ki}}}{1 + e^{\beta_0 + \sum \beta_k x_{ki}}} \quad (1)$$

The odds of crash occurrence $g(x)$ can be calculated as

$$g(x) = \frac{p_i}{1 - p_i} = e^{\beta_0 + \sum \beta_k x_{ki}} \quad (2)$$

where x_{ki} denotes input variables series; β_0 denotes the constant in the logistic model; β_k denotes coefficients for the independent input variables. Here β_k can be estimated by solving the log-likelihood function in

$$\ln L(\beta) = \sum_{i=1}^n [\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} - \ln(1 + e^{\beta_0 + \sum \beta_k x_{ki}})] \quad (3)$$

The crash risk model is estimated with the binary logistic regression procedure in SPSS 19. Backward LR (likelihood ratio) variables selection is applied to select the significant parameters in the proposed model. As shown in Table 2, results indicate that flow and speed dispersion are significant variables for crash risk estimation. The larger values of *Q* and *DV* indicate a higher crash risk. The *AUC* (area under the receiver operating characteristic curve, which illustrates the performance of the classifier) value indicates that the logistic regression model could successfully classify most of the crash and noncrash cases.

In order to compare the predicting ability of the HDMS data with the ability of the other kinds of traffic data, such as loop detector data [22] or Automatic Vehicle Identification (AVI) data [16], several existing studies are listed in Table 3. Table 3 reveals that the HDMS data provide better evidence for crash risk estimation and the corresponding model has

TABLE 4: Crash risk evaluation models for downstream and upstream of the HDMS.

Variables	<i>B</i>	<i>S.E.</i>	<i>Wals</i>	<i>df</i>	<i>Sig.</i>
<i>Q</i>	0.012	0.001	64.270	1	0.000
<i>DV</i>	0.139	0.026	27.961	1	0.000
<i>Constant</i>	-5.912	0.714	68.650	1	0.000
Downstream <i>AUC</i>	0.759				
Variables	<i>B</i>	<i>S.E.</i>	<i>Wals</i>	<i>df</i>	<i>Sig.</i>
<i>Q</i>	0.013	0.002	38.924	1	0.000
<i>DV</i>	0.093	0.029	10.320	1	0.001
<i>Constant</i>	-4.948	0.776	40.608	1	0.000
Upstream <i>AUC</i>	0.755				

a relatively better prediction accuracy despite of the simple form of model and the discrete distribution of the HDMS devices.

The spatial issue should be addressed for the implementation of VSL. Hence, another two models have been formulated to investigate the spatial effect, a downstream model and an upstream model. As shown in Figure 4, samples with crashes located downstream of the HDMS are the downstream samples and vice versa. Finally the downstream dataset contains 105 crashes and 360 noncrashes, and the upstream dataset contains 68 crashes and 266 noncrashes (the samples located just at the HDMS station are classified into both downstream and upstream samples).

As before, binary logistic regression has been used to estimate the crash risk models. The results are shown in Table 4.

Results indicate that the performance of the crash risk models considering the spatial effects is similar to the performance of the crash risk model for the whole segment. The reason for this is that the crash risk is stable on each segment and the traffic parameters of adjacent locations on the same segment are highly correlated, which has been shown by Fang et al. [25]. Hence, the model for the whole segment is used subsequently to estimate the real-time crash risk.

3. VSL Based on Microsimulation

3.1. Aimsun API. In order to verify the method based on dynamic VSL control of crash risk, a sub-segment of the G15 Freeway segment utilized in Section 2 is employed in the simulation (Segment 1 in Figure 2). The segment extends from “Nantong Toll Interchange” to “Sutong Bridge” and the total length is approximately 6 km.

Aimsun API (Application Programming Interface) can be a helpful platform to evaluate certain traffic management strategies. We can obtain the necessary real-time traffic data (flow, speed, occupancy, etc.) with required aggregation levels or detailed vehicle information. The project is built with Visual C++ 6.0 based on Visual Studio 2005. Using Aimsun API functions, the detectors, VMS, and traffic control plans are modeled and the attributes are defined in our in-depth simulations.

To simulate the real HDMS data, AIMSUN API function is used to gather the real-time vehicle information. The

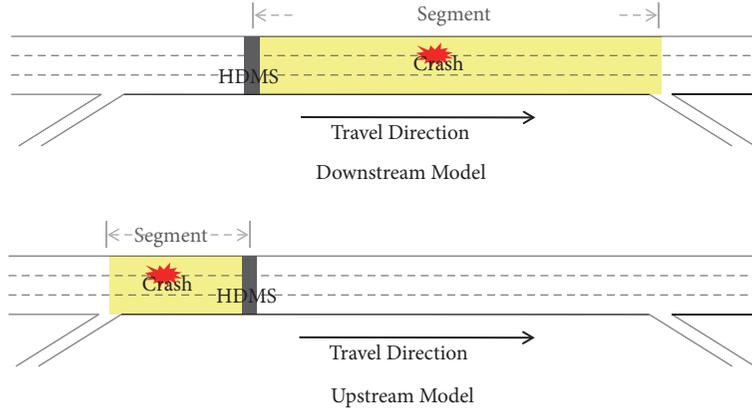


FIGURE 4: Spatial classification for downstream and upstream models.

scenarios are tested on the G15 Freeway with a design speed of 120km/h. In order to code the freeway segment in the simulation, Baidu Map GIS data source is utilized to build the freeway network. The drivers are assumed to comply with the speed limit, with a certain compliance rate, when the VSL starts to function on the segment.

The step size is 1 second. The Aimsun software development kit has been utilized to develop a module to extract the parameters for the crash risk evaluation model during the simulation process. The values for Q_i , DV_i , with a 5 minutes aggregation are recorded every minute. Thus the crash risk could be estimated every minute and exported as a report for final analysis as well as the average delay of all vehicles in the whole simulation process.

The real-time crash risk probability (p_i) for time period (i) is estimated with (4) using the parameters in Table 2:

$$\hat{p}_i = \frac{e^{-5.487+0.012*Q_i+0.12*DV_i}}{1 + e^{-5.487+0.012*Q_i+0.12*DV_i}} \quad (4)$$

The average delay (AD) for each simulation application can be calculated with

$$AD = \frac{\sum_{i=1}^n (T_i - T_{ei})}{n} \quad (5)$$

where T_i denotes the real time vehicle i spent on the freeway, T_{ei} denotes the expected time vehicle i spent on the freeway, and n denotes the number of vehicles passing by.

3.2. Simulation Calibration

3.2.1. Speed Distribution and Compliance Calibration. The speed distribution and compliance level are calibrated before the simulations. The original speed limit for the G15 Freeway is 120 km/h and the proportion of vehicles with speed above 120 km/h is set as the non-compliance level. Drivers tend to speeding on the freeway as the freeway is designed with better alignments especially long stretch of straight line. The traffic on the freeway mainly comprises private cars and trucks. All the speed data for May 2016 is used to calibrate the parameters. Figure 5 present the speed distributions for different vehicle types. The calibration includes the speed

TABLE 5: Simulation scenario settings in Aimsun.

Scenario Parameters	Settings
Road type	Freeway
Lane width	3.75 meter
Maximum speed limit	120 km/h
Detection cycle	1 second
Car following model	Minimum Headway+ deceleration estimation
Global arrivals	Exponential

distribution and the compliance rate. The compliance rate for cars and trucks is 88.37% and 79.05% respectively. Other parameters such as lane width, lane number and road type are also calibrated. Part of the final scenario settings are listed in Table 5.

3.2.2. Calibration of Traffic Temporal Distribution. Traffic spatial distribution should also be addressed to validate the simulations.

In existing studies, aggregate statistics have been validated such as the GEH statistics by FHWA [7]. In this study, HDMS data have been utilized to calibrate the simulations. 1-hour HDMS data starting at 2016-05-05, 13:50:00, are used to validate the simulation. The distribution diagram of time headway in Figure 6 indicates that the temporal distribution of the real data is similar to the temporal distribution of the simulation data, especially when the value of time headway exceeds 10 seconds. The significant difference appears on the distribution when the time headway is less than 4 seconds. The reason is that in real world some drivers may drive aggressively with a relatively low following distance especially under congestions, or try to change the lane with a relatively low transverse distance, leading to the relatively high proportion of low time headway. While in the simulation scenarios, the vehicles follow the car-following and lane-changing rules and they would never exceed the limit values. The result of an additional Pearson test shows that the coefficient value of the two curves is 0.75, which suggests moderate fidelity of the simulation.

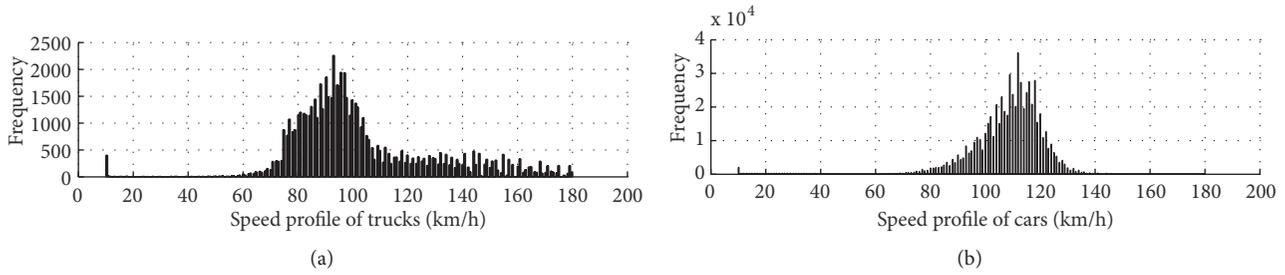


FIGURE 5: Speed distributions of vehicles (May 2016).

TABLE 6: Speed limit impacts under different flow conditions.

Flow condition (veh/h)	Speed Limit (VSL) (km/h)	Acceptable crash risk	Average Delay (s)	Crash risk benefit (%)	Average Delay Loss (%)
5000	60	0.23	153.24	29.81	258.98
4000	70	0.27	158.92	39.73	167.51
3000	80	0.16	44.64	25.47	35.89
2000	120	0.09	17.94	0	0

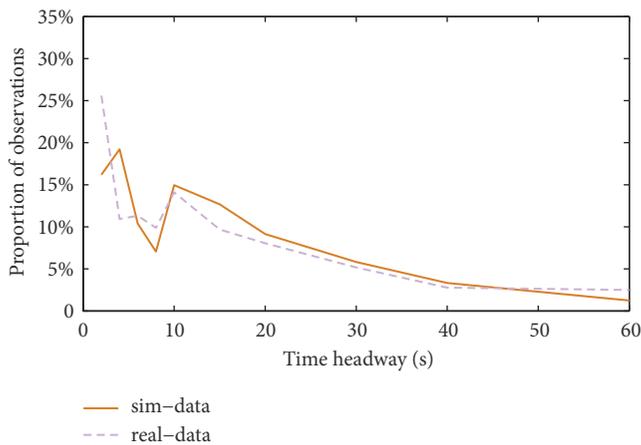


FIGURE 6: Comparison of the distributions of time headway.

3.3. VSL Performance Assessment under Different Flow Conditions. A set of six speed limits has been tested to evaluate the VSL performance under different flow conditions, namely 90, 80, 70, 60, 50, and 40 km/h. The traffic demand ranges from 2,000 to 5,000 veh/h. The Aimsun simulation results depend on the random seeds, reflecting the impact of random factors, and simulations were replicated five times to account for the variability. Each replication has 20 minutes to warm up with the traffic demand 2,000 veh/h and 60 minutes more to simulate the whole process with a different flow.

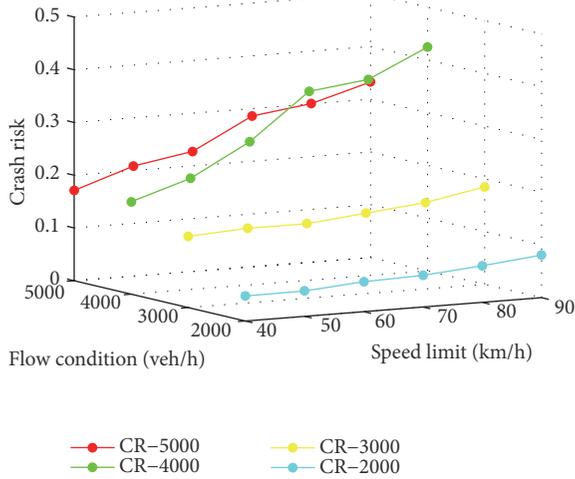
As the traffic is dynamic, the real traffic flow varies over time, as does the crash risk. Thus the objective is to keep the crash risk within an acceptable limit. In this study, the commonly used 85th percentile index in traffic safety is selected as the crash risk threshold for each replication; i.e., the traffic is evaluated as safe below that threshold. Once the crash risk exceeds the limit, proper strategies should be

implemented to minimize the risk. As inappropriate speed limits would decrease the capacity and increase traffic delay, a comprehensive analysis should be made to achieve an optimal cost benefit ratio. Figure 7 shows plots of crash risk (the possibility of crash occurrence) and average delay in relation to speed limit under different flow conditions.

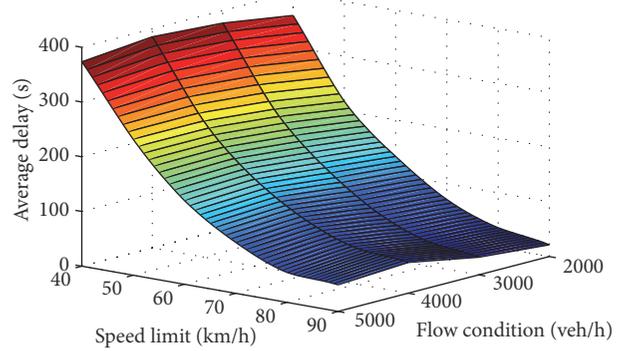
Figure 7 indicates that a lower speed limit would achieve lower crash risk; however the benefits vary at different flow levels. The average delay increases more and more as the speed limit is set lower. Hence, the crash risk benefit and average delay loss under different speed limits can be related to the original crash risk and the average delay of the replication without VSL, so that trade-offs can be made as suggested in Table 6. The results above are based on the hypothesis that all drivers would comply with the VSL; i.e., compliance rate equals 100%. However, when the VSL is advisory and also inappropriate, most of the drivers would not comply with it; thus the VSL strategy does not have any significant impact on traffic conditions [10]. Hence, simulations with a lower compliance rate are also examined, the rate being set to 50%. The results are shown in Figure 8. Figure 8 indicates that at flows of 2,000 to 3,000 veh/h and speed limits below 70 km/h, with a compliance rate of 50% the crash risk increases as the speed limit decreases. The compliance rate of 50% is low, with half of the drivers driving at the VSL and the other half driving as they prefer. At low flow levels, drivers would pursue higher driving speeds, so that speed dispersion would increase and make accidents more likely.

4. Application of VSL Strategies in Simulation

4.1. VSL Control Strategy. The objective of the VSL control strategy is to manage the traffic within an acceptable crash risk level and feedback is needed to adapt the strategy to the real-time traffic condition. Two kinds of strategies have been implemented in the simulations. The first is implementing



(a) Crash risk versus speed limits and flow conditions



(b) Average delay versus speed limits and flow conditions

FIGURE 7: Average delay and crash risk versus speed limit under different flow conditions.

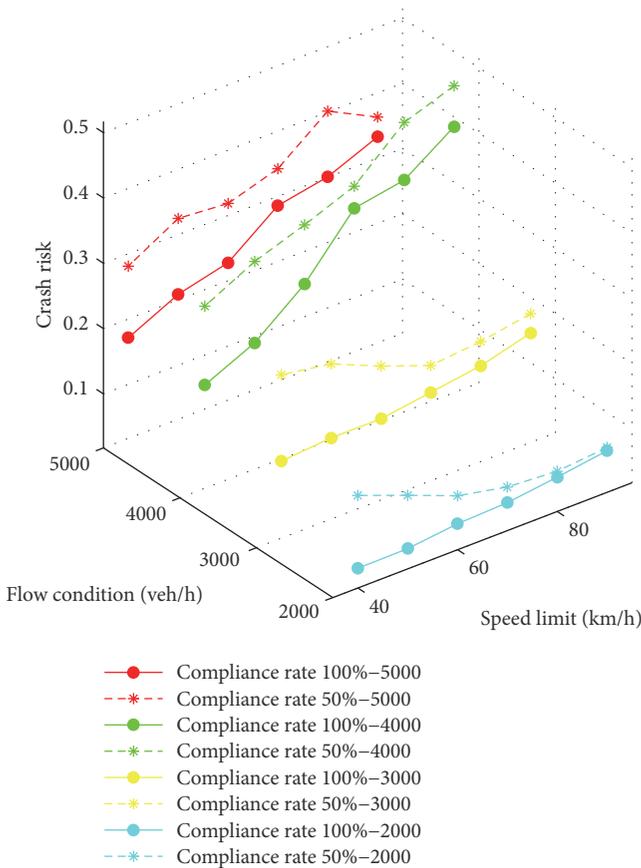


FIGURE 8: Comparison of crash risk versus speed limit for alternative compliance rates under different flow conditions.

and withdrawing the optimal VSL gradually (Strategy A) and the other is implementing and withdrawing the optimal VSL rapidly (Strategy B). Strategy A can be described as in Figure 9. When the real-time crash risk exceeds the

acceptable risk threshold τ), the corresponding target VSL V_τ is set and the VSL value is decreased at the gradient of 5 km/h every minute towards the target VSL V_τ . In Strategy B, when the crash risk exceeds τ , the VSL is set to V_τ immediately. The HDMS data are extracted from the HDMS database starting at 2016-05-02, 13:50:00. The corresponding traffic demand and speed distribution in the simulation are calibrated again to fit the real HDMS data. The compliance rates of the simulations is set at 88.37% for cars and 79.05% for trucks, being the statistical results in Section 3.2. It took 15 minutes to warm up and the whole simulation time is 75 minutes. The simulations, and the comparison with raw data, are labeled as “raw,” “Strategy A,” and “Strategy B.” The simulations are started by the same random seed.

4.2. Application Results. The simulations results with different strategies are shown in Figure 10. The crash risk curve of the raw HDMS data displays two crash risk peaks after 40 minutes, exceeding the crash risk threshold significantly. Compared with the raw curve, Strategy B could decrease the average crash risk by 10.15%, but the risk is still high after the simulation time of 50 minutes. The crash risk curve indicates that there are no significant crash risk peaks under the control of Strategy A. The average crash risk has been decreased by 22.63% by Strategy A compared with the Raw strategy curve. The crash risk remains at a low level. Strategy A outperforms Strategy B in which the speed limit is set to the target speed immediately. When drivers pass by the speed limit sign, they have to decelerate rapidly to comply with the speed limit. As a result, the speed dispersion increases rapidly and the crash risk increases as well.

Average delays are 56.31s, 24.87s, and 91.84s for Raw, Strategy A, and Strategy B, respectively. Thus Strategy A generates the shortest travel time and this strategy could control the traffic condition efficiently and steadily, whereas improper speed limit implementation may lead to unexpected traffic congestions.

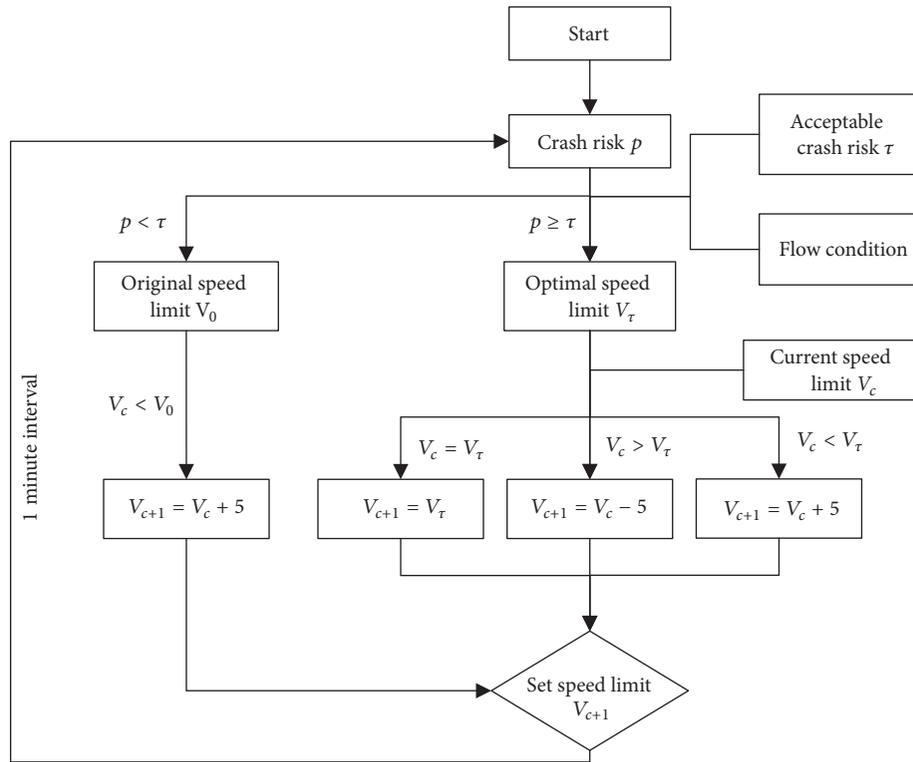


FIGURE 9: Strategy A: the VSL strategy with gradient feedback control.

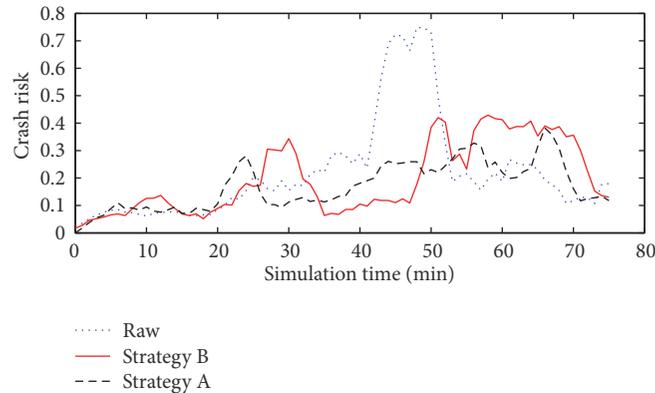


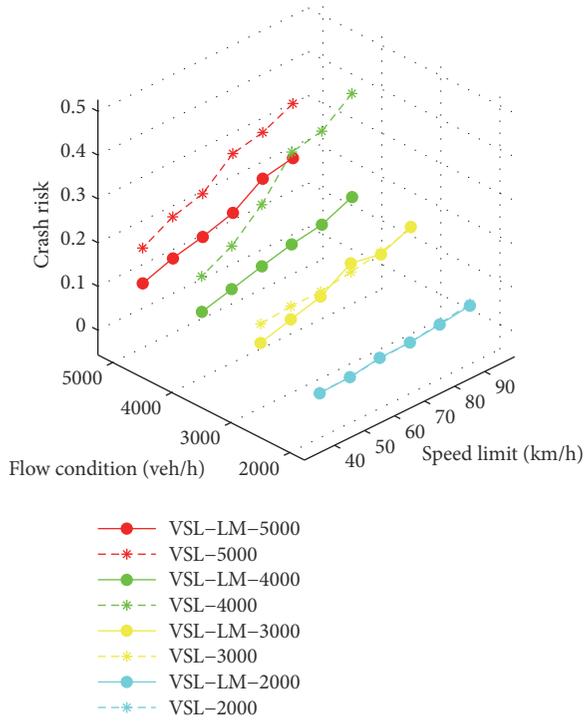
FIGURE 10: Simulation results under three conditions.

5. Discussion

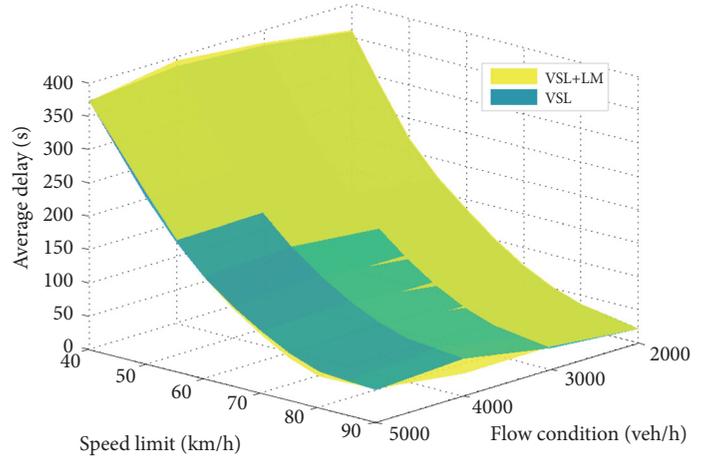
Results of this study demonstrate that the proposed VSL method could improve traffic safety, but more developments are required to produce integrated control strategies that are efficient and also applicable in real-time to large-scale networks [26]. Integrated control with ramp metering and VSL has been used to improve traffic flow efficiency based on an optimal VSL rate [27]. Lane management (LM) method has been employed in most countries, such as US, Europe, and China. Lord et al. [28] indicated that truck-free free-ways would have a better safety record than mixed traffic and separating truck traffic from passenger cars improves

safety. Toll lanes have been suggested to separate *Lights* and *Heavies* and the method could reduce total travel costs [29]. In this study, lane management has also been simulated. To separate the cars and the trucks, a solid line rule has been applied to restrict the lane-changing behavior on the mainline. The truck can only access the outer lane while the cars can access the other two lanes. Each condition with lane management has been simulated with different VSL under different traffic conditions. The results are shown in Figure 11.

Figure 11 indicates that, under the traffic conditions of 2,000 veh/h and 3,000 veh/h, lane management has little impact on the crash risk. However, under traffic conditions of



(a) Crash risk versus flow conditions and speed limits



(b) Average delay versus flow conditions and speed limits

FIGURE 11: Impact of lane management on crash risk and average delay under different traffic conditions.

4,000 veh/h or 5,000 veh/h, lane management has significant impacts on ameliorating the crash risk. By implementing lane management, the crash risk could be substantially controlled despite high speed limits.

As shown in Figure 11(b), lane management at lower traffic levels would have no impact on the average delay. Thus, by implementing lane management properly with VSL, traffic management administration would get significant improvement in crash risk without affecting the level of service of the freeway.

6. Conclusions

The study proposes an innovative dynamic variable speed approach through intervening in traffic flow dynamics. A binary logistic regression model based on HDMS data is built to estimate crash risk. HDMS data provide detailed vehicle information instead of aggregated data from loop detectors or other detectors. They provide better evidence on the crash mechanism. Microsimulations have been conducted with the AIMSUN simulation software. AIMSUN API is utilized to extract the detailed real-time vehicle information to calculate the crash risk. Different speed limits with several compliance rates under certain traffic conditions have been simulated. Considering the trade-off between safety benefits and travel time delay, we aim to optimize speed limit strategies under various traffic conditions.

Two kinds of VSL strategies have been applied to control the real-time crash risk in the simulated conditions of real

traffic accidents; the strategy of implementing and withdrawing the optimal VSL gradually (gradient control) could provide better control effects and keep the crash risk at a lower level. Furthermore, lane management control has also been assessed. Results indicate that such integrated control could significantly reduce the crash risk without increasing average traffic delay. The trends in optimal integrated traffic control to reduce real-time crash risk prove to be promising.

Several potential directions are open for future exploration. For example, further work is being conducted to study the performance of applying the strategies into various road types or road network. In future studies, as more and more surveillance devices and vehicle on-board devices are installed, real-time data such as weather condition as well as driving behavior could be obtained. Meanwhile, with the continuous spatial distribution of surveillance devices and detectors, the aggregated traffic control of multiple segments could be investigated to achieve balanced traffic conditions in the road network as more driver-friendly integrated control strategies are developed to fit the new era of ITS.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research is supported by the National Key Research and Development Program (2016YFC0802701), the National Natural Science Foundation (71301119, 71871161, and 51708421), Shanghai Pujiang Program, and the Fundamental Research Funds for the Central Universities. The authors are indebted to Professor Xiaobo Qu for his insightful suggestions to improve this manuscript.

References

- [1] A. Pande and M. Abdel-Aty, "Comprehensive analysis of the relationship between real-time traffic surveillance data and rear-end crashes on freeways," *Transportation Research Record*, no. 1953, pp. 31–40, 2006.
- [2] C. Xu, P. Liu, W. Wang, and Z. Li, "Evaluation of the impacts of traffic states on crash risks on freeways," *Accident Analysis & Prevention*, vol. 47, no. 1, pp. 162–171, 2012.
- [3] A. Theofilatos and G. Yannis, "A review of the effect of traffic and weather characteristics on road safety," *Accident Analysis & Prevention*, vol. 72, pp. 244–256, 2014.
- [4] X. Yang, Y. Lin, Y. Lu, and N. Zou, "Optimal variable speed limit control for real-time freeway congestions," *Procedia: Social and Behavioral Sciences*, vol. 96, pp. 2362–2372, 2013.
- [5] W. Zhizhou, L. Hao, and Y. Hongsu, "Variable speed limit control on highway work zone considering large vehicle mix rates," *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 1784–1791, 2013.
- [6] A. Tympakianaki, A. Spiliopoulou, A. Kouvelas, I. Papamichail, M. Papageorgiou, and Y. Wang, "Real-time merging traffic control for throughput maximization at motorway work zones," *Transportation Research Part C: Emerging Technologies*, vol. 44, pp. 242–252, 2014.
- [7] R. Yu and M. Abdel-Aty, "An optimal variable speed limits system to ameliorate traffic safety risk," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 235–246, 2014.
- [8] A. Hegyi, B. De Schutter, and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed limits," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 185–209, 2005.
- [9] S. Long, L. Gentry, and G. H. Bham, "Driver perceptions and sources of user dissatisfaction in the implementation of variable speed limit systems," *Transport Policy*, vol. 23, pp. 1–7, 2012.
- [10] A. Nissan and H. N. Koutsopoulos, "Evaluation of the impact of advisory variable speed limits on motorway capacity and level of service," *Procedia - Social and Behavioral Sciences*, vol. 16, pp. 100–109, 2011.
- [11] F. Soriguera, I. Martínez, M. Sala, and M. Menéndez, "Effects of low speed limits on freeway traffic flow," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 257–274, 2017.
- [12] M. Abdel-Aty, J. Dilmore, and A. Dhindsa, "Evaluation of variable speed limits for real-time freeway safety improvement," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 335–345, 2006.
- [13] X. Qu, S. Wang, and J. Zhang, "On the fundamental diagram for freeway traffic: a novel calibration approach for single-regime models," *Transportation Research Part B: Methodological*, vol. 73, pp. 91–102, 2015.
- [14] Y. Kuang, X. Qu, and S. Wang, "A tree-structured crash surrogate measure for freeways," *Accident Analysis & Prevention*, vol. 77, pp. 137–148, 2015.
- [15] Z. Li, S. Ahn, K. Chung, D. R. Ragland, W. Wang, and J. W. Yu, "Surrogate safety measure for evaluating rear-end collision risk related to kinematic waves near freeway recurrent bottlenecks," *Accident Analysis & Prevention*, vol. 64, pp. 52–61, 2014.
- [16] M. Ahmed, M. Abdel-Aty, and R. Yu, "Bayesian updating approach for real-time safety evaluation with automatic vehicle identification data," *Transportation Research Record*, no. 2280, pp. 60–67, 2012.
- [17] J. Zhang, X. Qu, and S. Wang, "Reproducible generation of experimental data sample for calibrating traffic flow fundamental diagram," *Transportation Research Part A: Policy and Practice*, vol. 111, pp. 41–52, 2018.
- [18] H. D. Son, Y.-J. Kweon, and B. B. Park, "Development of crash prediction models with individual vehicular data," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1353–1363, 2011.
- [19] E. I. Vlahogianni, M. G. Karlaftis, and F. P. Orfanou, "Modeling the effects of weather and traffic on the risk of secondary incidents," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 16, no. 3, pp. 109–117, 2012.
- [20] C. Xu, A. P. Tarko, W. Wang, and P. Liu, "Predicting crash likelihood and severity on freeways with real-time loop detector data," *Accident Analysis & Prevention*, vol. 57, pp. 30–39, 2013.
- [21] C. Lee, F. Saccomanno, and B. Hellinga, "Analysis of crash precursors on instrumented freeways," *Transportation Research Record*, no. 1784, pp. 1–8, 2002.
- [22] H.-C. Kwak and S. Kho, "Predicting crash risk and identifying crash precursors on Korean expressways using loop detector data," *Accident Analysis & Prevention*, vol. 88, pp. 9–19, 2016.
- [23] Y. Lao, G. Zhang, Y. Wang, and J. Milton, "Generalized nonlinear models for rear-end crash risk analysis," *Accident Analysis & Prevention*, vol. 62, pp. 9–16, 2014.
- [24] Z. Christoforou, S. Cohen, and M. G. Karlaftis, "Identifying crash type propensity using real-time traffic data on freeways," *Journal of Safety Research*, vol. 42, no. 1, pp. 43–50, 2011.
- [25] S. Fang, W. Xie, J. Wang, and D. R. Ragland, "Utilizing the eigenvectors of freeway loop data spatiotemporal schematic for real time crash prediction," *Accident Analysis & Prevention*, vol. 94, pp. 59–64, 2016.
- [26] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2043–2065, 2003.
- [27] R. C. Carlson, I. Papamichail, M. Papageorgiou, and A. Messmer, "Optimal motorway traffic flow control involving variable speed limits and ramp metering," *Transportation Science*, vol. 44, no. 2, pp. 238–253, 2010.
- [28] D. Lord, D. Middleton, and J. Whitacre, "Does separating trucks from other traffic improve overall safety?" *Transportation Research Record*, no. 1922, pp. 156–166, 2005.
- [29] A. de Palma, M. Kilani, and R. Lindsey, "The merits of separating cars and trucks," *Journal of Urban Economics*, vol. 64, no. 2, pp. 340–361, 2008.

Research Article

Bike-Sharing Static Rebalancing by Considering the Collection of Bicycles in Need of Repair

Sheng Zhang , Guanhua Xiang, and Zhongxiang Huang 

School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha 410114, China

Correspondence should be addressed to Zhongxiang Huang; mehzx@126.com

Received 16 May 2018; Revised 23 July 2018; Accepted 28 August 2018; Published 13 September 2018

Academic Editor: Zhiyuan Liu

Copyright © 2018 Sheng 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.

Bike-sharing systems, which are used in many cities worldwide, need to maintain a balance between the availability of bicycles and the availability of unoccupied bicycle slots. This paper presents an investigation of the net flow of each bike-sharing station in Jersey City. The data was recorded at 1-minute intervals. The sum of the initial bicycle number and the minimum net flow value was determined to be the demand for static rebalancing, and this led to the proposal of a bike-sharing demand prediction method based on autoregressive integrated moving average models. Considering that the existence of bicycles in a state of disrepair may adversely affect demand prediction and routine planning, we present an integer linear programming formulation to model bike-sharing static rebalancing. The proposed formulation takes into account the problem introduced by the need to collect bicycles in need of repair. A hybrid Discrete Particle Swarm Optimization (DPSO) algorithm was proposed to solve the model, which incorporates a reduced variable neighborhood search (RVNS) functionality together with DPSO to improve the global optimization performance. The effectiveness of the algorithms was verified by a detailed numerical example.

1. Introduction

Bike-sharing systems are designed to solve “the first and the last mile problem” and to provide a connection between multitransit modes. The most recent decade has witnessed the rapid worldwide development of bicycles as a transport mode driven by bike-sharing companies, such as Mobike and OFO. In this arrangement, the residents can rent or return bicycles at any public area, and this has greatly enhanced the convenience in the residents’ daily commutes. These systems have since spread all over the world. In contrast, the rapid development of bike-sharing has resulted in many problems, such as damaged bicycles, bike parking needs, and bicycle maintenance. In many cities, piles of bicycles are parked along public areas, and broken bicycles tend to be left carelessly. Among these violations, the parking problem is especially critical and has aroused people’s concern. These phenomena occur mainly because the operation management is not standardized and there is no effective rebalancing method.

Scholars and policy makers are committed to determining the causes and solution of these problems. Reinforcing the standardization of the bike-sharing system management

is essential; however, more importantly, advanced methods should be used to rebalance the stations to their optimum levels of occupation to avoid the inefficiency due to full and/or empty stations. Today’s competitive environment also requires bike-sharing companies to increase the operational efficiency by adopting such measures.

Bike-sharing systems are composed of several bike stations located at different sites across the city, and each station operates a number of bike slots from which bicycles can be collected or to which they can be returned. In a balanced situation, each bike station is required to have a certain number of empty slots, to allow arrivals, and a certain number of full slots, to allow departures. A certain amount of time after the beginning of the service, the users will have moved the bicycles among the stations of the system and the state will have deviated from the balanced one. Thus, these companies engage in the redistribution of bicycles, an operation known as rebalancing. Rebalancing, which is performed by vehicles with the necessary capacity, is normally required at the end of the day, when the system is closed or when the use of the system can be considered negligible. In this case, the rebalancing is considered to be static. Some bike-sharing

operators may require the rebalancing to be performed when the system is open and in operation, incurring a situation known as dynamic rebalancing. In this paper, we present the static case, which is referred to as the Static Bike-sharing Rebalancing Problem (SBRP).

As discussed elsewhere [1–4], SBRP is a vehicle route problem, modeled on a complete digraph $G = (V, A)$, where $V = \{0, 1, \dots, n\}$ is the set of vertices, including the depot (vertex 0) and n stations (vertices 1, . . . , n), and A is the set of arcs between each pair of vertices. The traveling cost c_{ij} is associated with the arc $(i, j) \in A$. For each vertex $i \in V$, a request q_i is given, with $q_0 = 0$. Requests can be either positive or negative. If $q_i \geq 0$, then i is a pickup station, from which q_i bicycles should be removed; if $q_i < 0$, then i is a delivery station, to which q_i bicycles should be supplied, for $i \in V \setminus \{0\}$. Bicycles supplied to delivery stations can either come from the depot or from pickup stations. Vehicles do not necessarily have to leave the depot empty. The objective is to operate a fleet of m identical vehicles of capacity Q that need to be available at the depot to respond to requests and to minimize the total cost of the traversed arcs.

However, the existence of bicycles in disrepair would disrupt the demand prediction and rebalancing efforts. In reality, after a long period of usage, some of the bicycles would require repairs. If the bicycles in need of repair continue to remain in a station indefinitely, they may be mistaken for operational bicycles by the system operator who would not deliver new bicycles to that station. Meanwhile, the customer's demand at that station cannot be satisfied. Thus, the vehicles would need to collect these bicycles in disrepair and return them to the depot, thereby increasing the complexity of the static rebalancing problem. Contrary to the traditional vehicle routing problem with simultaneous pickup and delivery (VRPSPD), the number of bicycles in need of repair collected from some stations cannot be used to respond to requests of delivery stations. As these bicycles are expected to take up a particular amount of capacity of a vehicle, the path of the vehicle would need to be newly planned and rescheduled.

This study makes the following substantial contributions to the bike-sharing static rebalancing problem:

- (1) We conducted a pattern analysis of each bike-sharing station over the data set of Jersey City. On the basis of high-resolution net flow data analysis, we found that there is always a fixed input/output pattern in each bicycle station. Consequently, the bicycle stations can be divided into two types: demand-station and supply-station. We also found that the bicycle demand is determined by the minimum net flow value. For a demand-station, the bicycle demand is equal to the current number of bicycles plus the minimum net flow value; for supply-station, the supply of station is equal to the current number of bicycles plus the total difference (net flow value at the end of the day). Therefore, we established the new demand prediction equations based on bike-sharing station pattern analysis. These equations are easy to implement in practice.

- (2) To the best of our knowledge, our work is the first to formulate the bike-sharing static rebalancing problem by considering the need to collect bicycles requiring repair. The existence of these bicycles increases the complexity of the static rebalancing problem. The number of bicycles requiring repair collected from some stations cannot be used to respond to requests of delivery stations. The bicycles in disrepair will take up a particular amount of capacity of a vehicle. We build new integer linear programming formulations for the bike-sharing static rebalancing problem by considering the need to collect bicycles requiring repair. The presented model is expected to ensure the specific constraint that the sum of requests of the visited stations and the bicycles in disrepair plus the initial load is never negative or greater than the capacity along the route traveled by a vehicle.

The remainder of this paper is organized as follows. Previous related research is reviewed in Section 2. In Section 3 we briefly present the real-world case study conducted in Jersey City, by analyzing the travel flows, and propose the station classification model based on state curve analysis. We then propose a demand prediction method based on the station classification method and autoregressive integrated moving average models. In Section 4, we consider the problem presented by the need to collect bicycles in disrepair and formulate a new bike-sharing static rebalancing model. In Section 5, we propose the hybrid algorithm of DPSO-VNS to solve the proposed model. In Section 6, a case study is presented to show the solution procedure. Conclusions and future work are presented in Section 7.

2. Literature Review

Lately, the bike-sharing rebalancing problem has received increased attention from the practitioners as well as researchers because of its great importance in the bike-sharing system operation. Many researchers viewed the SBRP as a special Traveling Salesman Problem (TSP) [3], multiple-TSP [5], Pickup and Delivery Traveling Salesman Problem (PDTSP) [1, 2, 6–11], and Vehicle Routing Problem with Simultaneous Pickup and Delivery (VRPSPD) [12], or built mixed integer linear programming (MILP) formulations with different objections or constrains [13–18].

Table 1 summarizes the static BRP (Bike Rebalancing Problem) publications according to the formulations, problem objectives, solutions, determination of demand, and whether the collection of damaged bicycles is considered in the formulations. Only a few studies have presented the bicycle demand determination method. To the best of our knowledge, only Alvarez-Valdes et al. [2] studied the return of damaged bicycles, but the impact of damaged bicycles to the vehicle's capacity constraint has not been considered in the published literature. There are three main solution methods, e.g., exact method [1, 3, 9], heuristics method [2, 10, 12, 13, 15–17], and hybrids of these two methods [8, 14, 18].

To solve a problem such as the VRPSPD, the general method is to establish a mixed integer linear programming

TABLE 1: Summary of the characteristics of the static BRP in the literature.

References	Bicycle demand	Formulations	Objective	Damaged bicycles	Solution
Benchimol et al. J. H. Lin et al.	Given	Single-vehicle 1-PDTSP	Minimize total travel cost	Without Consideration	Approximation algorithm
Raviv et al.	A decision variable	Extends the 1-PDTSP	Minimize the weighted sum of total travel time and penalty cost	Without Consideration	A two-stage heuristic
Chemla et al. Dell'Amico et al.	Given	Extends the 1-PDTSP	Minimize total travel cost	Without consideration	Branch-and-cut algorithm
Ho and Szeto.	A decision variable	A single-vehicle MIP	Minimize total penalty cost	Without consideration	Iterated tabu search heuristic
Angeloudis et al.	Determined by the history net flow data	m-TSP	Minimize the total duration of travel	Without Consideration.	Approximation algorithm
Forma et al.	A decision variable	3-step mathematical programming-based heuristic	Minimize total cost, consisting of the sum of penalties and the operating costs.	Without Consideration.	3-step math heuristic
Alvarez-Valdes et al.	A decision variable	Multi-commodity PDTSP	Minimize the weighted sum of the total service time and the coefficient of variations of the duration of all routes	Yes, but the impact of the damaged bicycles to the vehicle's capacity constraint is not considered	A heuristic algorithm based on minimum cost flow problem and an insertion algorithm
Szeto et al.	Given	Single-vehicle 1-PDTSP	Minimize the weighted sum of unmet customer demand and operational time on the vehicle route.	Without Consideration	Enhanced Chemical Reaction Optimization
Li et al.	Given	MIP	Minimize total cost	Without Consideration.	A combined hybrid genetic algorithm
Kadri et al.	Given	TSP	Minimize the total waiting time of the stations	Without consideration	Branch and bound algorithm
Cruz et al.	Given	1-PDTSP	Minimize the traveling cost	Without Consideration.	Iterated Local Research-based heuristic.
Ho and Szeto.	Given	1-VRPSPD	Minimize the sum of penalty cost	Without Consideration.	A hybrid large neighborhood search combined with a tabu search.
Schuijbroek et al.	Determine service level requirements at each station	A Clustered MIP	Minimize maximum tour length	Without consideration	Cluster-first route-second heuristic
Szeto and Shui.	Given	MIP	Minimize the total demand dissatisfaction (TDD) and total service time	Without consideration	An enhanced artificial bee colony algorithm
Bulhões et al.	Given	MIP	Minimize the total travel time	Without consideration	Branch-and-cut algorithm combined with local search metaheuristic
This study	Present a predict method using the operating data	Multi-commodity VRPSPD	Minimize the traveling cost	Yes, the vehicle's capacity constraint to the damaged bicycles is considered	A hybrid DPSO-VNS

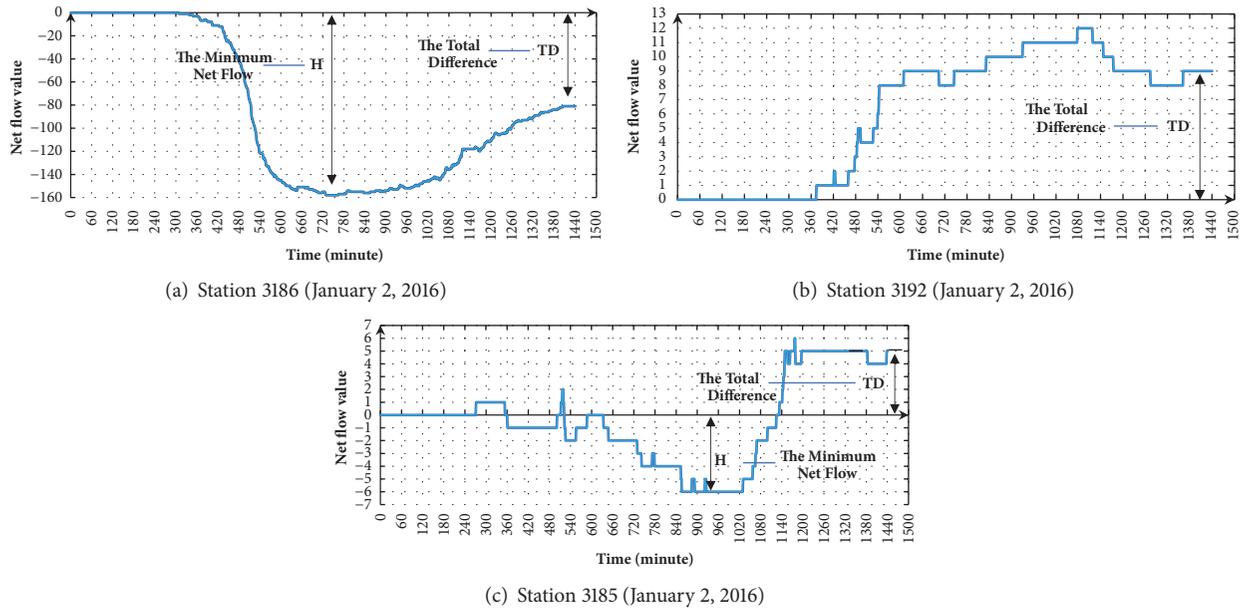


FIGURE 1: Net flow of typical stations.

(MILP) formulation with different objectives or constrains. Our model for the bike-sharing static problem extends the multicommodity VRSPD [19, 20]. In the proposed model, the vehicles are required not only to deliver goods to customer, but also to pick up goods at customer location [21].

Currently, there is no general algorithm to find the global optimal solution for this kind of nonconvex and nonlinear programming problems. Heuristic algorithm can get high-quality solutions, with good computational efficiency. We have searched the literature which uses heuristic and metaheuristic approaches solving such problems, such as Ant Colony System algorithm, D-Ants, and local search algorithm. Obviously, the static rebalancing problem is also a combinational optimization problem, and the solution space is a discrete space made up of a limited number of integral points. Hence, Discrete Particle Swarm Optimization (DPSO) is applicable to solve this kind of problem. In this paper, we adopt a hybridization algorithm combining DPSO and Variable Neighborhood Search (VNS) to solve the VRSPD, which is proposed by F. P. Goksal et al. [21] in order to enhance the global search ability of DPSO and overcome the possibility of entrapment in local optimum and make full use of the local search ability of variable neighborhood search.

3. Demand Prediction

3.1. Pattern Analysis of the State of a Station. Most existing methods to predict the short-time demand of the stations are based on the historical pattern of the bicycle number change at each station. Andreas et al. [22] detected the temporal and geographic mobility patterns based on the data from the operator's website. These patterns were applied to the prediction of the available bicycle number at any station.

After a fairly long period since a bicycle sharing system operated at an area, the resident will develop a stationary trip habit gradually. And the trip habits will tend to be stable.

There will be a state periodic similarity at each bike-sharing station. That is to say, there is a fixed input/output pattern in bicycle station. We utilize operational data to analyze and find this input/output pattern. On the basis of the detailed data analysis we find that there are mainly two types of bicycle station: demand and supply-station, and the bicycle demand is determined by the minimum net flow value. For a demand-station and a demand-supply-station, the bicycle demand is equal to the current number of bicycles plus the minimum net flow value; for supply-station, the supply of station is equal to the current number of bicycles plus the total difference (net flow value at the end of the day).

One of the advantages is that forecasting the bicycle demand through analyzing the pattern of each station is more precise than other methods, such as traffic mode split forecast. Currently, the operational data is adequate. With the accumulation of operational data, the demand pattern recognition of each bicycle station will be more accurate. At the same time, the ARIMA mode we utilized is a method that has been widely used as a famous linear time series model, which usually has the advantage of high accuracy, efficiency, and adapting ability.

The bike-sharing system used for the case study is located in Jersey City, which has only 51 stations as shown in Figure 1. This bike-sharing system operates during the daytime but remains closed at night, which is quite common in small/medium cities. The redistribution of bicycles is carried out during the night by means of a single vehicle that visits each station exactly once. The twelve-month historical operation data is used for the case study, which records the details of the trips including the departure and arrival time of each trip.

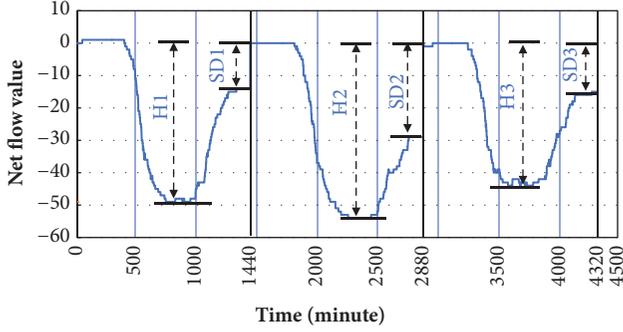


FIGURE 2: State curve of station 3186 from January 2 to January 4, 2016.

The station state is characterized by the net flow on one-minute basis, which is calculated as the difference between the inflow and the outflow. From the net flow curve of each station, we can determine the difference between the bicycles available at the beginning of the day and those remaining at the end of the day. For convenience of observation, we set the initial net flow to zero. Figure 1 shows the typical net flow curves of three stations, where the x- and y-axes represent the times series over one day and the difference between arrivals and departures from a station per minute, respectively.

Based on the state curve, the stations can be divided into three categories, namely, demand-station, supply-station, and demand-supply-station. These different types of stations play different roles in the bike-sharing system. For the demand-station, the number of the outgoing bicycles always exceeds the number of incoming bicycles, such as station 3186 shown in Figure 1(a). For the supply-station, the number of outgoing bicycles is smaller than the number of incoming bicycles. Station 3129 shown in Figure 1(b) is an example of the supply-station, whose net flow remains positive, which indicates that the bicycles of this station can be delivered to other stations. At the third type of stations, demand-supply-station, the net flow can be both positive and negative, such as station 3185 shown in Figure 1(c).

3.2. The Demand Equation. The bicycle demand is the number of bikes required in stations, which can be defined based on the station state curve. As shown in Figure 2, the real maximum demand of a day is determined by the minimum net flow value denoted by H_t , which is illustrated by the vertical line in Figure 2. We use TD_t to denote the total difference between the bicycle number at the end of a day and that at the beginning of a day, which reflects the imbalance of the bicycle arrival and departure. Let N_t be the number of bicycles at the end of the previous day. We can easily find that the bicycle demand of a station q_t is equal to $N_t + H_t$.

Based on similar analysis, we can define the bicycle demand of three types of the stations as follows.

(1) *Demand-Station and Demand-Supply-Station.* For a demand-station and a demand-supply-station, $H_{i,t}$ is negative. The bicycle demand is calculated by

$$q_{i,t} = N_{i,t} + H_{i,t} \quad (1)$$

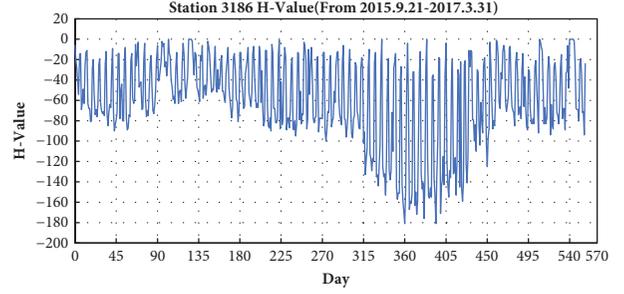


FIGURE 3: Maximum demand ($H_{i,t}$) curve of station 3186 from Sep. 21, 2015 to Jan. 31, 2017.

where $q_{i,t}$ is the demand of station i of the t th day; $H_{i,t}$ is the minimum demand of station i on the t th day (to be predicted); and $N_{i,t}$ is the number of bicycles at station i at the end of the $(t - 1)$ th day.

(2) *Supply-Station.* For supply-station, $H_{j,t}$ is zero and $TD_{j,t}$ is positive. The bicycle demand of the supply-station is defined as

$$q_{j,t} = N_{j,t} + TD_{j,t} \quad (2)$$

where $q_{j,t}$ is the supply of station j on the t th day; $TD_{j,t}$ is the total difference between the bicycle number at the end of the t th day and that at the beginning of the that day of station j ; and $N_{j,t}$ is the number of bicycles at station j at the end of the $(t - 1)$ th day.

(3) *Shortage of the Bike-Sharing System.* For a bike-sharing system, the total bicycle demand is equal to $\sum_i q_{i,t}$, and total supply is $\sum_j q_{j,t}$. If the total demand is greater than the total supply, then the bike-sharing system needs more bicycles. The number of bicycles that need to be supplied is equal to $\sum_i q_{i,t} - \sum_j q_{j,t}$.

If the total demand is smaller than the total supply, the bike-sharing system does not need new supplies, and the redundant bicycles do not need to be sent to the depot.

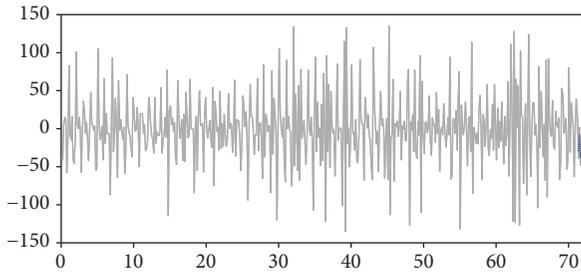
3.3. Maximum Daily Demand Prediction Model. In (1) and (2), $N_{i,t}$ and $N_{j,t}$ are given and $H_{i,t}$ and $TD_{j,t}$ need to be predicted. There are many methods that can be used to handle the time series, such as multiple linear regression [23], exponential smoothing [24], artificial neural networks (ANN) [25], and methods based on a multimodel fusion algorithm [26]. Among the abovementioned methods, autoregressive integrated moving average (ARIMA) model is the most straightforward method to predict $H_{i,t}$ and $TD_{j,t}$ value. ARIMA method has been widely used as a famous liner time series model, which usually has the advantage of high accuracy, efficiency, and adapting ability [27, 28].

The following shows the prediction process of $H_{i,t}$. One can predict $TD_{j,t}$ by using the same method, so these will not be reviewed here.

Figure 3 shows the maximum demand ($H_{i,t}$) curve of station 3186 from Sep. 21, 2015, to Jan. 31, 2017.

TABLE 2: Prediction error analysis of the $H_{i,t}$ value.

Date Weekday	Feb. 1, 2017 Sunday	Feb. 2, 2017 Monday	Feb. 3, 2017 Tuesday	Feb. 4, 2017 Wednesday	Feb. 5, 2017 Thursday	Feb. 6, 2017 Friday	Feb. 7, 2017 Saturday
Predicted value y	-1	-29	-20	-27	-28	-42	-16
Measured value x	-2	-24	-19	-35	-33	-33	-18
Absolute Difference $ x_i - y_i $	1	5	1	8	5	9	2
Absolute Percentage Errors	50%	20%	5%	22%	15%	27%	11%

FIGURE 4: Prediction results for the $H_{i,t}$ value of Mondays by using ARIMA(3,1,3).

Because the demand pattern is related to the day of a week, we divided them into categories based on the day of the week. Figure 4 shows the prediction results of the $H_{i,t}$ value of the Monday series. After calibration, we find that the ARIMA(3,1,3) model is the best for Monday series. The forecasting result is shown in Table 2. Due to the fact that the real demand value in station 3186 is small, the absolute percent errors actually indicate a small absolute difference, so this result is acceptable for a bike-sharing system.

With the proposed method, we have already the demand of each station. We can obtain the number of bicycles in need of repair at each station by checking and user report, which are used to guide the rebalancing work.

4. Formulation of the VRSPD Problem

To solve a problem like Vehicle Routing Problem with Simultaneous Picked and Delivery (VRSPD), the common method is building a mixed integer linear programming (MILP) formulations with different objections or constrains. An integer linear programming formulation is proposed in this paper to model bike-sharing static rebalancing by considering the collection of bicycles in need of repair. The formulation is derived from Min (1989) [20] and Dethloff (2001) [19]. We enhanced the existing SBRP methodology in the following three ways: (a) the number of vehicles that perform the task may be greater than one; (b) the number of items picked up or delivered to each customer is predicted by a dedicated method derived from the net flow pattern analysis; and (c) there are two types of commodity in the model, ready bicycles, and damaged bicycles. The impact of damaged bicycles on the vehicle's capacity constraint is considered in the proposed model. In the proposed model, the vehicles

are required not only to deliver goods to customer, but also to pick up goods at customer location. This is imposed by constraints (11)-(16) in the paper.

Specifically, the static rebalancing problem should involve determining how to drive at most m vehicles through the graph, with the aim of minimizing the total cost and ensuring that the following constraints are not violated: (i) Each vehicle performs a route that starts and ends at the depot. (ii) Each vehicle starts from the depot empty or with some initial load (i.e., with a number of bicycles that vary from 0 to Q). (iii) Each station is visited exactly once and its request is completely fulfilled as a result of this visit. (iv) The number of bicycles in need of repair at each station p_i is given, and the quantities collected at pickup stations can be used to respond to the requests of delivery stations or can be returned directly to the depot; however, bicycles in disrepair cannot be used to meet the requests of delivery stations and need to be returned to the depot. (v) Finally, the sum of requests of the visited stations and the bicycles in disrepair plus the initial load is never negative or greater than Q along the route followed by a vehicle.

Notations

V : set of vertices, where $V_0 \subseteq V \setminus \{0\}$, $S \neq \emptyset$

A : set of arcs

n : number of stations

p_i : bicycles in disrepair at vertex i , $p_i \geq 0$

K : set of vehicles

k : the k th vehicle, where $k = 1, 2, \dots, m$

m : number of vehicles

Q : vehicle capacity

q_i : demand at vertex i

c_{ij} : cost of the arc (i, j)

L_{0k} : initial load of the k th vehicle

L_{k0} : final load of the k th vehicle when it returns to the depot

L_{kj} : load of the k th vehicle after vertex j , $j \in V_0$

f_{ij} : flow of arc (i, j) , that is the load of the vehicle passing arc (i, j) , if any, for $(i, j) \in A$

Q_{tot} : the total demand

g_{ij} : the space remaining in the vehicle after the arc (i, j) , if any, for $(i, j) \in \tilde{A}$

x_{ijk} : a binary variable.

The formulation for bike-sharing static rebalancing by considering the collection of bicycles in need of repair is given as follows:

$$\min \sum_{i \in V} \sum_{j \in V} \sum_{k \in K} c_{ij} x_{ijk} \quad (3)$$

subject to

$$\sum_{i \in V} \sum_{k \in K} x_{ijk} = 1 \quad \forall j \in V_0 \quad (4)$$

$$\sum_{i \in V} x_{ihk} = \sum_{j \in V} x_{hjk} \quad \forall h \in V_0; \forall k \in K \quad (5)$$

$$\sum_{j \in V_0} \sum_{k \in K} x_{0jk} \leq m \quad (6)$$

$$\sum_{j \in V_0} x_{0jk} = \sum_{i \in V_0} x_{i0k} \quad \forall k \in K \quad (7)$$

$$Q_{tot} = \sum_{j \in V_0} q_j \quad (8)$$

$$L_{0k} = \max \left\{ 0, - \sum_{i \in V} \sum_{j \in V_0} q_j x_{ijk} \right\} \quad \forall k \in K \quad (9)$$

$$\sum_{k \in K} L_{0k} \geq \max \{0, -Q_{tot}\} \quad (10)$$

$$L_{k0} = \sum_{i \in V} \sum_{j \in V_0} p_j x_{ijk} \quad \forall k \in K \quad (11)$$

$$\sum_{k \in K} L_{k0} = \sum_{j \in V_0} p_j \quad (12)$$

$$L_{jk} \geq L_{0k} + q_j + p_j - M(1 - x_{0jk}) \quad (13)$$

$\forall k \in K; \forall j \in V_0$

$$L_{kj} \geq \theta_{ki} + q_j + p_j - M \left(1 - \sum_{k \in K} x_{ijk} \right) \quad (14)$$

$\forall i \in V_0; \forall j \in V_0; i \neq j$

$$L_{0k} \leq Q, \quad \forall k \in K \quad (15)$$

$$L_{kj} \leq Q, \quad j \in V_0, k \in K \quad (16)$$

$$\sum_{i \in V_0} \sum_{j \in V_0} \sum_{k \in K} x_{ijk} \leq |S| - \left\lceil \frac{\max \{0, -Q_{tot}\} + \sum_{i \in V_0} p_i}{Q} \right\rceil, \quad (17)$$

$$x_{ijk} = \begin{cases} 1 & \text{if the arc } (i, j) \text{ was used by vehicle } k \\ 0 & \text{otherwise.} \end{cases} \quad (18)$$

Objective function (3) minimizes the traveling cost. Constraints (4) require that every node except the depot is visited

exactly once. Constraints (6) and (7) ensure, respectively, that at most m vehicles leave the depot, and that all vehicles that are used return to the depot after completing their route. Constraints (9) impose the initial load of the k th vehicle. The total load leaving the initial depot should be nonnegative in any case; moreover, in case Q_{tot} takes a negative value, it should not be lower than this value. This fact is imposed by constraints (10).

Constraints (11) impose the final load of the k th vehicle. Constraints (12) state that the total load entering the depot is the number of all collected bicycles in need of repair. Constraints (13) impose the vehicle load after the first station. Constraints (14) impose the vehicle load after station j . Constraints (15) and (16) specify the upper bounds on the loads.

Constraints (17) state that, for each subset S of vertices, the number of arcs with both their tail and head in S should not exceed the cardinality of S minus the minimum number of vehicles required to serve S . An estimation of the minimum number of vehicles is simply obtained by computing the value of the sum of the demands and the sum of the bicycles in need of repair, dividing it by the vehicle capacity and then rounding up the result.

5. A Hybrid Algorithm Based on DPSO and VNS

It is intractable to use exact methods such as those shown in Table 1 to solve large, realistic rebalancing problems, because the proposed problem is an NP-hard problem [4]. Previous studies [8, 13] have also illustrated this point by conducting numerical experiments. Hence, most of the existing studies focused on heuristic and metaheuristic approaches that can produce high-quality solutions within limited computational times [7, 8, 29, 30]. Ai and Kachitvichyanukul proposed Particle Swarm Optimization (PSO) algorithms for the VRPSPD [31]. Gajpal and Abad proposed an Ant Colony System (ACS) algorithm to solve the VRPSPD. However, the ACS is more complicated than other intelligence algorithms, such as PSO algorithms [32]. Berbeglia et al. suggested that the problem of VRPSPD falls under the category of a one-to-one pickup and delivery problem [29]. Reimann et al. proposed a decomposition approach (D-Ants) for the VRP model [33]. Li et al. discovered an Iterated Local Search algorithm, which is based on a metaheuristic. Their results showed that this algorithm is more accurate than others. However, the algorithm is computationally less efficient [34]. Liu et al. applied an algorithm for solving the Dynamic VRPSPD model based on POS of a Hopfield Neural Network. This algorithm combined the ability of the Hopfield Neural Network to reach the nearest point from the initial minimum point towards the descendent direction of gratitude and the advantage of the global search capability of the POS algorithm [35].

Obviously, the static rebalancing problem is a combinatorial optimization problem, and the solution space is a discrete space made up of a limited number of integral points. Hence, Discrete Particle Swarm Optimization (DPSO) is applicable to solve this kind of problem. Compared with a

Genetic algorithm, there is no mutation rate and cross rate parameters in DPSO; compared with the Simulated Annealing Method, there is no cumbersome process involving temperature drop; compared with the Ant Colony algorithm, there is no variety of complex judgment steps and parameter setting. In addition, the formula used to update the speed and position is used to ensure that all the particles converge to the global optimal value. Thus, the DPSO has a strong global searching capability.

However, there are some drawbacks in the DPSO algorithm. First, setting the different parameters will have an obvious effect on the DPSO performance. Second, the DPSO algorithm may easily converge to a local optimum. Hence, to overcome these shortcomings while retaining the advantages, a new hybridization algorithm combining DPSO and Variable Neighborhood Search (VNS) was proposed by Goksal et al. [21]. This algorithm can enhance the global search ability of DPSO, overcome the possibility of entrapment in local optimum, and make full use of the local search ability of variable neighborhood search.

The hybrid algorithm combined DPSO and VNS employed here includes four steps: representation of the solution, initializing the discrete particles, updating the positions of particles, and the strengthened local search based on the VNS algorithm.

5.1. Representation of the Solution. Finding a suitable representation of the solution is important. A good representation of the solution can enhance the performance of the DPSO algorithm and decrease the complexity of the algorithm. In this study we used the following procedures.

There are N bicycle stations in the bicycle system, and the request is that the maximum load per route does not exceed the vehicle capacity (Q) and the remaining vehicle volume is sufficiently large to accommodate the bicycles from the next station after collection from or delivery to the last station. Each particle is a solution which contains several routes. The total number of particles is initialized with $N = \{1, 2, 3, \dots, N_0\}$ positive integer numbers. Then each particle inserts zero (standing for the depot) under the constraints of the capacity of the vehicle and the space remaining in the vehicle. For instance, $X = [1, 3, 6, 5, 0, 8, 11, 2, 4, 0, 7, 10, 9]$ is a solution, for which the particle has three routes. Route 1 is 0-1-3-6-5-0, route 2 is 0-8-11-2-4-0, and route 3 is 0-7-10-9-0. The cost of each particle is calculated by

$$\text{cost} = C_{0,X(i)} + \sum_{k=i+1}^{j-1} C_{X(k),X(k+1)} + C_{X(j),0} \quad (19)$$

where $C_{i,j}$ is the cost in arc (i, j) , $X_{(k)}$ is the number of X particles representing the bicycle stations, and zero is the depot.

5.2. Initialization of the Discrete Particle Swarm. Commonly, the initial particle swarm is often a random product and we also adapt this format. Particles have different dimensions for different problems. Thus, the revised algorithm for gaining an initial particle swarm is provided in Algorithm 1.

5.3. Position and Velocity Updating Rule for Particles. Generally, the velocity and position of the DPSO entail four basic operations, which are briefly mentioned here. The velocity and position updating rule of a particle is based on the following two equations:

(1) The velocity of a particle at iteration t is calculated by using four parts, the position minus, the position multiply, the velocity multiply, and the velocity and position, as shown below.

$$V_{ij}^{t+1} = wV_{ij}^t \oplus c_1r_1 (P_{ij}^t - X_{ij}^t) \oplus c_2r_2 (G_{ij}^t - X_{ij}^t) \quad (20)$$

(2) The position of a particle at iteration t is updated by considering the operation of position plus velocity. The position updating rule is given by

$$X_{ij}^{t+1} = X_{ij}^t \oplus V_{ij}^t \quad (21)$$

where \oplus is a newly defined operation.

5.4. Local Search Based on VNS Algorithm. As has been proposed in the literature review, there are many neighborhood structures for us to use. In our implementation, we utilize three neighborhood structures, i.e., insert (i, j) , swap (i, j) , and $2 - \text{Opt}(i, j)$. These neighborhood structures are briefly introduced as follows:

Insert(i,j): This structure enables us to randomly produce two positive integer numbers (i and j) limited to the range of the particle, which indicate, respectively, the number of bicycles collected by customers and the position at which the particle is inserted and that both of them can belong to a different route; then the randomly produced number is collected and inserted into a randomly produced position.

Swap(i,j): This structure indicates that a customer i is removed from route 1 and a customer j is removed from route 2. Then these chosen customers $cs1$ and $cs2$ are reinserted into different routes, route 2 and route 1, respectively. Moreover, the route and the customer number cannot be adjacent but can be noncontiguous.

2-Opt(i,j): This structure reverses the direction of a path lying between the customers $i-1$ and j by replacing nonadjacent arcs $(i-1, j)$ and $(j-1, j)$ belonging to the same route with $(i-1, j-1)$ and (i, j) , respectively. Equation (22) is used to update the personal best.

$$P_i^t = \begin{cases} P_i^{t,E} & \text{if } f(P_i^{t,E}) < f(P_i^{t,o}) \\ P_i^{t,o} & \text{if } f(P_i^{t,E}) \geq f(P_i^{t,o}) \end{cases} \quad (22)$$

where P_i^t is the final personal best particle, $P_i^{t,E}$ is the updated personal best particle operated by the three neighborhood structures, and $P_i^{t,o}$ is the original personal best particle without any neighborhood structure operating. The cost function is expressed by (19).

Figure 5 illustrates the application of these neighborhood structures, and the VNS algorithm is shown in Algorithm 2. The hybridization of DPSO with VNS is shown in Algorithm 3.

```

Input: The number of bike stations, the cost matrix and the supply and demand per station
Output: The initial solution (also known as the initial particle)
Begin
  Initialize the discrete particle swarm based on the population parameter and the dimensional
  of particles, known as  $X^t$ 
  Loading the base data i.e., the cost matrix and supply and demand per station
  While  $n < N$  do
    Begin
    Load  $\leftarrow 0$ 
    Route  $\leftarrow$  null
     $r \leftarrow 0$ 
    for  $i = 1, 2, \dots, N$  do
      repeat
        Load = Load +  $P_i + Q_i$  ( $P_i$  is the supply account,  $Q_i$  is the demand account)
        Free Space = Capacity - Load (Capacity is the capacity of the vehicle)
        if Load > Capacity or Free Space <  $P_i$ 
           $O_j = i - 1$ 
          end
          for  $i = 1, 2, \dots, N$  do
            repeat
               $r \leftarrow r + 1; k \leftarrow O_j$ 
              for  $m = k + 1$  to  $j$  do
                Route( $r$ ) =  $X(i) \cup$  Route( $r$ )
              end
            end
          end
         $n \leftarrow n + 1$ 
      end
    Return Route
  end

```

ALGORITHM 1: The initial particle swarm algorithm.

```

Input: Initial vehicle routes
Output: Improved vehicle routes by VNS algorithm
Begin
  Initialize sequence of three neighborhood structures (1-insert, 2-swap, 3-2-opt) and randomly
  produce a positive integer of no more than 3
  Choose the  $k^{th}$  neighbor ( $S'$ ) of  $S' \in N(S)$  from these three neighborhood structures and apply
   $k^{th}$  neighborhood to  $P$  and obtain ( $P'$ )
  if  $F(P') < F(P)$  then  $P \leftarrow P', F(P) \leftarrow F(P')$ 
    and apply the remainder of the neighborhood structures in sequence one by one to ( $P$ ) and
    obtain ( $P'$ )
    if  $F(P') < F(P)$  then  $P \leftarrow P', F(P) \leftarrow F(P')$ 
  else
    delete the  $k^{th}$  neighborhood structure and apply the remainder of the neighborhood
    structures one by one to ( $P$ ) and obtain ( $P'$ )
  return  $P$  and  $F(P)$ 
end

```

ALGORITHM 2: Local search algorithm (VNS).

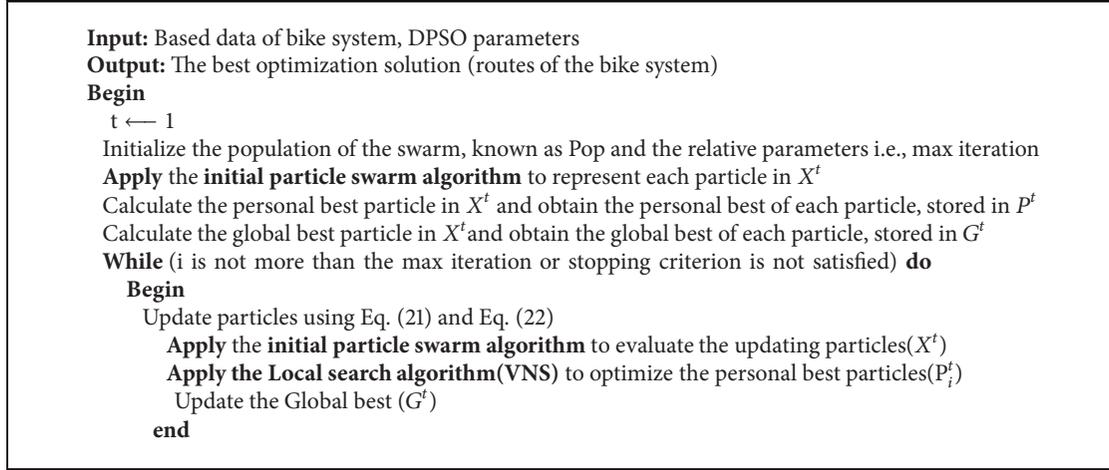
6. Case Study

We assume a fictitious small-scale bicycle system as an illustrative example to show the efficiency of the proposed model and solving method. The system contains 14 bike stations. The daily supply and demand, the number of bicycles in need of repair, and the adjacency matrix for distance

serving as cost matrix are given, separately, in Tables 3, 4, and 6.

The hybrid algorithm based on DPSO and VNS for rebalancing bike sharing has the following process.

6.1. Generating the Initial Discrete Particle Swarm. Normally, the initial discrete particle swarm is produced randomly,



ALGORITHM 3: Hybridization of DPSO with VNS.

TABLE 3: Daily supply and demand.

Station-ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Demand and Supply	12	-2	8	5	-7	4	-2	-8	2	1	1	-3	3	10

TABLE 4: Number of bicycles in disrepair.

Station-ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Bicycles in disrepair	1	5	4	2	2	4	3	2	1	0	3	2	4	7

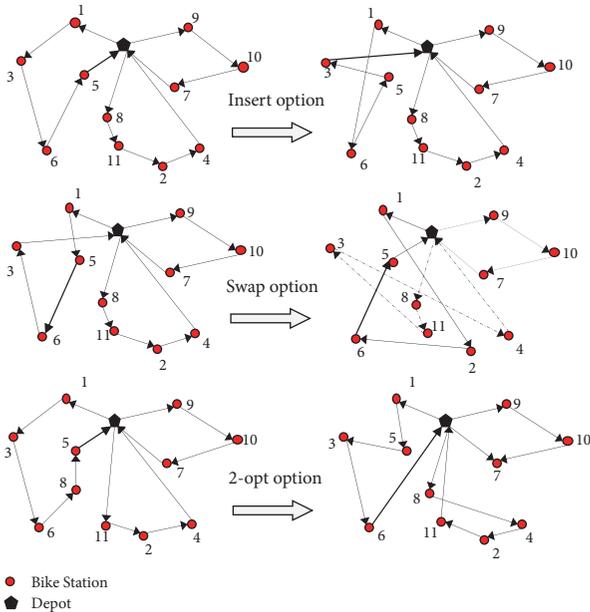


FIGURE 5: Three neighborhood structures.

but choosing an appropriate initial swarm can expedite the convergence rate of the algorithm. In order to gain the favorable population, we initialize the discrete particle swarm randomly, belonging to the arrangement $x_{ij}, x_{ij} \in [1, 20]$,

$i \in [1, popsize], j \in [1, length]$, and obtain the initial particle $X_i = \{x_{ij} \mid x_{ij} \in [1, 20], i \in [1, popsize], j \in [1, length]\}$.

According to the size of this fictitious bike-sharing system, we set 20 as the solution scale. Divide the particles according to the constraint condition, and obtain the pattern of particles that was revealed in Table 5. Numbers against asterisk mark denote the depot and the end of one route.

6.2. Fitness Evaluation. The feasible solutions for this system are produced in Section 6.1, so the calculation of these feasible solutions is the main topic of this section. We utilize the fitness value of each particle as an optimized target; the fitness function is a standard for screening particles; hence we choose the reciprocal of the objective function as the fitness function. The specific calculation formula is

$$Fitness(t) = c_{0, X_t(i)} + \sum_{k=i+1}^{j-1} c_{X_t(k), X_t(k+1)} + c_{X_t(j), 0} \quad (23)$$

where

$X_t(i)$ represents the value of the i th data of the t th particle,

$c_{i,j}$ represents the vehicle cost (distance) from station i to j ,

$fitness$ is the fitness value of the particle.

The following process of the first particle shows the steps to calculate the fitness value.

TABLE 5: Initial Particles.

1st	6	3	11	7	0*	14	8	5	0*	1	2	4	13	0*	9	10	12
2nd	7	1	13	2	0*	14	6	10	0*	12	11	4	0*	8	3	9	5
3rd	2	10	4	0*	5	3	8	0*	12	11	7	1	0*	13	14	6	9
4th	10	9	0*	6	13	2	1	0*	11	14	3	8	7	0*	4	5	12
5th	8	9	x	0*	12	2	10	3	0*	7	4	1	13	11	0*	5	6
6th	2	3	0*	11	1	6	14	0*	5	9	8	0*	4	13	12	7	10
7th	9	2	1	0*	6	11	0*	13	10	8	14	3	0*	7	4	12	5
8th	14	7	10	8	0*	2	6	12	0*	1	3	4	5	0*	11	9	13
9th	3	7	1	0*	13	9	14	8	0*	11	4	10	0*	5	6	12	2
10th	11	13	0*	1	5	14	12	0*	8	10	9	4	0*	3	7	6	2
11th	2	10	13	0*	7	9	0*	12	14	1	6	5	11	0*	8	3	4
12th	9	10	0*	5	1	11	6	7	0*	4	14	12	0*	13	8	2	3
13th	10	6	4	0*	2	7	14	0*	13	1	8	5	0*	3	12	9	11
14th	1	10	0*	7	2	3	9	13	0*	14	6	0*	5	8	4	12	11
15th	8	14	6	0*	1	7	9	4	0*	2	11	13	0*	10	3	5	12
16th	3	11	10	0*	4	7	12	0*	13	5	14	2	0*	6	8	9	1
17th	4	1	14	9	0*	2	10	7	0*	13	12	8	6	0*	3	5	11
18th	6	10	14	0*	11	12	4	3	0*	1	2	13	0*	9	7	8	5
19th	11	1	0*	10	8	14	0*	3	13	2	9	0*	5	12	4	7	6
20th	9	11	6	0*	7	10	14	0*	5	1	2	0*	13	4	3	8	12

TABLE 6: Cost Matrix.

Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.00	0.26	0.18	0.21	0.35	0.43	0.40	0.29	0.23	0.29	0.27	0.20	0.08	0.20	0.27
2	0.26	0.00	0.11	0.16	0.40	0.44	0.45	0.46	0.13	0.17	0.45	0.34	0.30	0.22	0.29
3	0.18	0.11	0.00	0.07	0.30	0.36	0.36	0.36	0.18	0.12	0.34	0.23	0.20	0.12	0.19
4	0.21	0.16	0.07	0.00	0.24	0.29	0.29	0.32	0.25	0.08	0.30	0.20	0.20	0.06	0.13
5	0.35	0.40	0.30	0.24	0.00	0.08	0.06	0.20	0.48	0.24	0.20	0.17	0.28	0.19	0.11
6	0.43	0.44	0.36	0.29	0.08	0.00	0.06	0.27	0.54	0.27	0.28	0.25	0.36	0.25	0.17
7	0.40	0.45	0.36	0.29	0.06	0.06	0.00	0.22	0.54	0.29	0.23	0.21	0.33	0.25	0.16
8	0.29	0.46	0.36	0.32	0.20	0.27	0.22	0.00	0.50	0.36	0.02	0.12	0.21	0.26	0.22
9	0.23	0.13	0.18	0.25	0.48	0.54	0.54	0.50	0.00	0.29	0.48	0.38	0.30	0.29	0.37
10	0.29	0.17	0.12	0.08	0.24	0.27	0.29	0.36	0.29	0.00	0.35	0.25	0.27	0.11	0.14
11	0.27	0.45	0.34	0.30	0.20	0.28	0.23	0.02	0.48	0.35	0.00	0.10	0.19	0.24	0.21
12	0.20	0.34	0.23	0.20	0.17	0.25	0.21	0.12	0.38	0.25	0.10	0.00	0.12	0.14	0.13
13	0.08	0.30	0.20	0.20	0.28	0.36	0.33	0.21	0.30	0.27	0.19	0.12	0.00	0.17	0.21
14	0.20	0.22	0.12	0.06	0.19	0.25	0.25	0.26	0.29	0.11	0.24	0.14	0.17	0.00	0.08
15	0.27	0.29	0.19	0.13	0.11	0.17	0.16	0.22	0.37	0.14	0.21	0.13	0.21	0.08	0.00

Note: Number 1 represents the depot.

1st 6 3 11 7 0* 14 8 5 0* 1 2 4 13 0* 9 10 12

Labeling method with list: Beginning from (0, 0) in the cost matrix, we obey these rules: Odd and even numbers of particles are indicated as rows and columns in the cost matrix, respectively. Further, when the adjacent number in the cost matrix is a cross, it is often the cost of the route. When all the numbers, exactly all the adjacent numbers, have been labeled,

this process is complete. Last, the consequence is equal to the sum of all the labeled points.

We present a representation of the cost matrix of the first particle in Table 7.

The fitness value is obtained by calculating the remaining particles in this way. The result is shown in Table 8. Numbers

TABLE 7: Cost calculation progress of routes 1, 2, 3, and 4.

Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0.00 [*]	0.26 ^③ [§]	0.18	0.21	0.35	0.43 ^② ^{&}	0.40 ^① [#]	0.29	0.23	0.29 ^④ [@]	0.27	0.20	0.08 ^④ [@]	0.20	0.27 ^② ^{&}
1	0.26	0.00	0.11	0.16	0.40	0.44	0.45	0.46	0.13	0.17	0.45	0.34	0.30	0.22	0.29
2	0.18	0.11 ^③ [§]	0.00	0.07	0.30 ^③ [§]	0.36	0.36	0.36	0.18	0.12	0.34	0.23	0.20	0.12	0.19
3	0.21	0.16	0.07	0.00	0.24	0.29	0.29 ^① [#]	0.32	0.25	0.08	0.30	0.20 ^① [#]	0.20	0.06	0.13
4	0.35	0.40	0.30	0.24	0.00	0.08	0.06	0.20	0.48	0.24	0.20	0.17	0.28	0.19	0.11
5	0.43	0.44	0.36	0.29	0.08	0.00	0.06	0.27	0.54	0.27	0.28	0.25	0.36	0.25	0.17
6	0.40	0.45	0.36	0.29	0.06	0.06	0.00	0.22	0.54	0.29	0.23	0.21	0.33	0.25	0.16
7	0.29 ^① [#]	0.46	0.36	0.32	0.20	0.27	0.22	0.00	0.50	0.36	0.02	0.12 ^① [#]	0.21	0.26	0.22
8	0.23	0.13	0.18	0.25	0.48	0.54 ^② ^{&}	0.54	0.50	0.00	0.29	0.48	0.38	0.30	0.29	0.37 ^② ^{&}
9	0.29	0.17	0.12	0.08	0.24	0.27	0.29	0.36	0.29	0.00	0.35	0.25	0.27	0.11	0.14
10	0.27	0.45	0.34	0.30	0.20	0.28	0.23	0.02	0.48	0.35 ^④ [@]	0.00	0.10	0.19 ^④ [@]	0.24	0.21
11	0.20	0.34	0.23	0.20	0.17	0.25	0.21	0.12	0.38	0.25	0.10	0.00	0.12	0.14	0.13
12	0.08	0.30	0.20	0.20	0.28	0.36	0.33	0.21	0.30	0.27	0.19	0.12	0.00	0.17	0.21
13	0.20 ^③ [§]	0.22	0.12	0.06	0.19 ^③ [§]	0.25	0.25	0.26	0.29	0.11	0.24	0.14	0.17	0.00	0.08
14	0.27	0.29	0.19	0.13	0.11	0.17	0.16	0.22	0.37	0.14	0.21	0.13	0.21	0.08	0.00

①[#] Route 1 ②[&] Route 2 ③[§] Route 3 ④[@] Route 4

TABLE 8: Fitness value.

	Initial Particles																fitness	
1st	6	3	11	7	0 [*]	14	8	5	0 [*]	1	2	4	13	0 [*]	9	10	12	4.89
2nd	7	1	13	2	0 [*]	14	6	10	0 [*]	12	11	4	0 [*]	8	3	9	5	4.70
3rd	2	10	4	0 [*]	5	3	8	0 [*]	12	11	7	1	0 [*]	13	14	6	9	4.58
4th	10	9	0 [*]	6	13	2	1	0 [*]	11	14	3	8	7	0 [*]	4	5	12	3.25
5th	8	9	x	0 [*]	12	2	10	3	0 [*]	7	4	1	13	11	0 [*]	5	6	4.21
6th	2	3	0 [*]	11	1	6	14	0 [*]	5	9	8	0 [*]	4	13	12	7	10	5.24
7th	9	2	1	0 [*]	6	11	0 [*]	13	10	8	14	3	0 [*]	7	4	12	5	6.35
8th	14	7	10	8	0 [*]	2	6	12	0 [*]	1	3	4	5	0 [*]	11	9	13	4.25
9th	3	7	1	0 [*]	13	9	14	8	0 [*]	11	4	10	0 [*]	5	6	12	2	4.68
10th	11	13	0 [*]	1	5	14	12	0 [*]	8	10	9	4	0 [*]	3	7	6	2	4.78
11th	2	10	13	0 [*]	7	9	0 [*]	12	14	1	6	5	11	0 [*]	8	3	4	5.21
12th	9	10	0 [*]	5	1	11	6	7	0 [*]	4	14	12	0 [*]	13	8	2	3	5.36
13th	10	6	4	0 [*]	2	7	14	0 [*]	13	1	8	5	0 [*]	3	12	9	11	4.25
14th	1	10	0 [*]	7	2	3	9	13	0 [*]	14	6	0 [*]	5	8	4	12	11	6.25
15th	8	14	6	0 [*]	1	7	9	4	0 [*]	2	11	13	0 [*]	10	3	5	12	5.48
16th	3	11	10	0 [*]	4	7	12	0 [*]	13	5	14	2	0 [*]	6	8	9	1	5.36
17th	4	1	14	9	0 [*]	2	10	7	0 [*]	13	12	8	6	0 [*]	3	5	11	7.21
18th	6	10	14	0 [*]	11	12	4	3	0 [*]	1	2	13	0 [*]	9	7	8	5	4.87
19th	11	1	0 [*]	10	8	14	0 [*]	3	13	2	9	0 [*]	5	12	4	7	6	4.98
20th	9	11	6	0 [*]	7	10	14	0 [*]	5	1	2	0 [*]	13	4	3	8	12	4.56

against asterisk mark denote the depot and the end of one route.

6.3. Calculation of the Historical and Global Optimal Values of the Particle. The historical and global optimal values are the particle's cognitive and social messages, which have a significant impact on the probability of finding the preferable particle. Hence, the definition for the historical optimal particle is as follows: particles with the best historical fitness

and global optimal particle; the best particles with the best historical fitness are mentioned. According to the definition, we obtain all the particles' historical optimal values and the particle swarm's global value at each iteration, as shown in Table 9.

In the first iteration, we initialize the original fitness value of the particles as the personal best and initialize the minimum of the personal best as the global best. The consequence is represented in Table 10.

TABLE 9: Historically optimal and globally optimal particles.

$x_{1,1}^t$	$x_{1,2}^t$	$x_{1,3}^t$...	$x_{1,length-2}^t$	$x_{1,length-1}^t$	$x_{1,length}^t$	f_1^t	1st Particle's Personal Best
$x_{2,1}^k$	$x_{2,2}^k$	$x_{2,3}^k$...	$x_{2,length-2}^k$	$x_{2,length-1}^k$	$x_{2,length}^k$	f_2^k	2nd Particle's Personal Best
...
$x_{20,1}^l$	$x_{20,2}^l$	$x_{20,3}^l$...	$x_{20,length-2}^l$	$x_{20,length-1}^l$	$x_{20,length}^l$	f_{20}^l	20th Particle's Personal Best
Global Best							$\min \{f_i^t \mid i \in [1, popsize], t \in [1, Maxiterate]\}$	

TABLE 10: Initial fitness value, personal best, and global best of the particles.

	Initial Particles															fitness	Personal best		
1st	6	3	11	7	0*	14	8	5	0*	1	2	4	13	0*	9	10	12	4.89	4.89
2nd	7	1	13	2	0*	14	6	10	0*	12	11	4	0*	8	3	9	5	4.70	4.70
3rd	2	10	4	0*	5	3	8	0*	12	11	7	1	0*	13	14	6	9	4.58	4.58
4th	10	9	0*	6	13	2	1	0*	11	14	3	8	7	0*	4	5	12	3.25	<u>3.25</u> (Global best)
5th	8	9	x	0*	12	2	10	3	0*	7	4	1	13	11	0*	5	6	4.21	4.21
6th	2	3	0*	11	1	6	14	0*	5	9	8	0*	4	13	12	7	10	5.24	5.24
7th	9	2	1	0*	6	11	0*	13	10	8	14	3	0*	7	4	12	5	6.35	6.35
8th	14	7	10	8	0*	2	6	12	0*	1	3	4	5	0*	11	9	13	4.25	4.25
9th	3	7	1	0*	13	9	14	8	0*	11	4	10	0*	5	6	12	2	4.68	4.68
10th	11	13	0*	1	5	14	12	0*	8	10	9	4	0*	3	7	6	2	4.78	4.78
11th	2	10	13	0*	7	9	0*	12	14	1	6	5	11	0*	8	3	4	5.21	5.21
12th	9	10	0*	5	1	11	6	7	0*	4	14	12	0*	13	8	2	3	5.36	5.36
13th	10	6	4	0*	2	7	14	0*	13	1	8	5	0*	3	12	9	11	4.25	4.25
14th	1	10	0*	7	2	3	9	13	0*	14	6	0*	5	8	4	12	11	6.25	6.25
15th	8	14	6	0*	1	7	9	4	0*	2	11	13	0*	10	3	5	12	5.48	5.48
16th	3	11	10	0*	4	7	12	0*	13	5	14	2	0*	6	8	9	1	5.36	5.36
17th	4	1	14	9	0*	2	10	7	0*	13	12	8	6	0*	3	5	11	7.21	7.21
18th	6	10	14	0*	11	12	4	3	0*	1	2	13	0*	9	7	8	5	4.87	4.87
19th	11	1	0*	10	8	14	0*	3	13	2	9	0*	5	12	4	7	6	4.98	4.98
20th	9	11	6	0*	7	10	14	0*	5	1	2	0*	13	4	3	8	12	4.56	4.56

Note: Numbers against asterisk mark denote the depot and the end of one route.

TABLE 11: Variable neighborhood operation.

$x_{1,1}^{t'}$	$x_{1,2}^{t'}$	$x_{1,3}^{t'}$...	$x_{1,length-2}^{t'}$	$x_{1,length-1}^{t'}$	$x_{1,length}^{t'}$	$f_1^{t'}$	1st Particle's Personal Best
$x_{2,1}^{k'}$	$x_{2,2}^{k'}$	$x_{2,3}^{k'}$...	$x_{2,length-2}^{k'}$	$x_{2,length-1}^{k'}$	$x_{2,length}^{k'}$	$f_2^{k'}$	2nd Particle's Personal Best
...
$x_{20,1}^{l'}$	$x_{20,2}^{l'}$	$x_{20,3}^{l'}$...	$x_{20,length-2}^{l'}$	$x_{20,length-1}^{l'}$	$x_{20,length}^{l'}$	$f_{20}^{l'}$	20th Particle's Personal Best
After variable neighborhood operation Global Best							$\min \{f_i^{t'} \mid i \in [1, popsize], t' \in [1, Maxiterate]\}$	

Note: $x_{i,j}^{k'}$ is the result of the variable neighborhood operation.

6.4. Local Search through the Variable Neighborhood Operation. Our algorithm utilizes the variable neighborhood search algorithm given in Table 11 to enhance the ability to choose the solution randomly, ensure diversity of the individual, and overcome the disadvantage that the solution is a local optimum. First, we updated the individual particles' historical best and global best. Then, in our implementation, three variable neighborhood structures are applied to the particle's personal best mentioned at Section 5.4.

Based on the particular variable neighborhood structure, we used the personal best as an operation by taking the

2-opt operation structure as an example. The steps and consequence are the following.

Procedure 1. Produce a group of 2-opt structures randomly; i.e., in the first particle, there are three points, (3, 7), (14, 5), and (2, 10), respectively.

Procedure 2. Subject the particle to the 2-opt operation. An example using the 1st particle is provided in Table 12.

The remaining particles were subjected to operation according to the form procedure, and the final result is presented in Table 13.

TABLE 12: Particles subjected to neighborhood structure operation (2-Opt).

1st	6	3	11	7	0*	14	8	5	0*	1	2	4	13	0*	9	10	12
After 2-opt-1st	6	3	7&	11&	0*	14	5&	8&	0*	1	2	10&	9&	0&	13&	4&	12
2nd	7	1	13	2	0*	14	6	10	0*	12	11	4	0*	8	3	9	5
3rd	2	10	4	0*	5	3	8	0*	12	11	7	1	0*	13	14	6	9
4th	10	9	0*	6	13	2	1	0*	11	14	3	8	7	0*	4	5	12
5th	8	9	14	0*	12	2	10	3	0*	7	4	1	13	11	0*	5	6
6th	2	3	0*	11	1	6	14	0*	5	9	8	0*	4	13	12	7	10
7th	9	2	1	0*	6	11	0*	13	10	8	14	3	0*	7	4	12	5
8th	14	7	10	8	0*	2	6	12	0*	1	3	4	5	0*	11	9	13
9th	3	7	1	0*	13	9	14	8	0*	11	4	10	0*	5	6	12	2
10th	11	13	0*	1	5	14	12	0*	8	10	9	4	0*	3	7	6	2
11th	2	10	13	0*	7	9	0*	12	14	1	6	5	11	0*	8	3	4
12th	9	10	0*	5	1	11	6	7	0*	4	14	12	0*	13	8	2	3
13th	10	6	4	0*	2	7	14	0*	13	1	8	5	0*	3	12	9	11
14th	1	10	0*	7	2	3	9	13	0*	14	6	0*	5	8	4	12	11
15th	8	14	6	0*	1	7	9	4	0*	2	11	13	0*	10	3	5	12
16th	3	11	10	0*	4	7	12	0*	13	5	14	2	0*	6	8	9	1
17th	4	1	14	9	0*	2	10	7	0*	13	12	8	6	0*	3	5	11
18th	6	10	14	0*	11	12	4	3	0*	1	2	13	0*	9	7	8	5
19th	11	1	0*	10	8	14	0*	3	13	2	9	0*	5	12	4	7	6
20th	9	11	6	0*	7	10	14	0*	5	1	2	0*	13	4	3	8	12

Note: Numbers against asterisk mark denote the depot and the end of one route. Numbers with & mark were subjected to operation by the 2-opt structure.

TABLE 13: Particles subjected to 2-Opt structure operation.

After 2-opt-1st	6	3	7&	11&	0*	14	5&	8&	0*	1	2	10&	9&	0&	13&	4&	12
After 2-opt-2nd	7	13&	1&	2	0*	14	4&	11&	12&	0&	10&	6&	0*	8	5&	9&	3&
After 2-opt-3rd	4&	10&	2&	0*	5	3	7&	11&	12&	0&	8&	1	0*	13	14	6	9
After 2-opt-4th	10	9	0*	6	1&	2&	13&	0*	11	14	3	8	5&	4&	0&	7&	12
After 2-opt-5th	8	14&	9&	0*	12	2	10	3	0*	7	4	11&	13&	1&	0*	5	6
After 2-opt-6th	2	3	0*	11	1	6	14	0*	5	9	4&	0&	8&	13	12	7	10
After 2-opt-7th	9	6&	0&	1&	2&	11	0*	13	10	8	14	4&	7&	0&	3&	12	5
After 2-opt-8th	14	7	10	2&	0&	8&	6	12	0*	5&	4&	3&	1&	0&	11	9	13
After 2-opt-9th	3	7	13&	0&	1&	9	14	8	0*	11	4	10	0*	5	6	12	2
After 2-opt-10th	11	13	0*	1	14&	5&	12	0*	8	10	9	2&	6&	7&	3&	0&	4&
After 2-opt-11th	2	0&	13&	10&	7	9	0*	12	14	1	6	5	11	0*	8	3	4
After 2-opt-12th	9	10	0*	1&	5&	11	6	4&	0&	7&	14	12	0*	13	8	2	3
After 2-opt-13th	10	2&	0&	4&	6&	7	14	0*	13	1	5&	8&	0*	3	12	9	11
After 2-opt-14th	1	10	7&	0&	2	3	9	13	14&	0&	6	0*	5	8	4	12	11
After 2-opt-15th	8	6&	14&	0*	1	7	9	4	0*	2	11	0&	13&	10	3	5	12
After 2-opt-16th	3	11	10	0*	4	7	12	0*	13	5	0&	2&	14&	6	8	9	1
After 2-opt-17th	4	9&	14&	1&	0*	2	10	7	0*	13	12	8	3&	0&	6&	5	11
After 2-opt-18th	6	10	11&	0&	14&	12	4	3	0*	1	2	13	7&	9&	0&	8	5
After 2-opt-19th	11	1	0*	10	8	9&	2&	13&	3&	0&	14&	0*	5	12	4	7	6
After 2-opt-20th	9	11	7&	0&	6&	10	14	0*	5	1	8&	3&	4&	13&	0&	2&	12

Note: Numbers against asterisk mark denote the depot and the end of one route. Numbers with & mark were subjected to operation by the 2-opt structure.

Then, according to the constraint of the model, we check and recode particles. If the operated particles can satisfy the constraint, we accept these particles; otherwise, we recode the operated particles. The result of checking and recoding can be seen in Table 14.

6.5. Particle Updating Rules and Algorithm Termination Condition. The velocity and position of a particle are updated according to (21) and (22), after which iterative operation occurs. The particles are updated according to the relative operation relating to the position and velocity.

TABLE 14: Recoding particles after the 2-Opt operation.

After 2-opt-1st	6	3	7	0*	11	14	5	8	0*	1	2	10	9	0*	13	4	12
After 2-opt-2nd	7	13	0*	1	2	14	4	0*	11	12	10	6	0*	8	5	9	3
After 2-opt-3rd	4	10	2	0*	5	3	7	11	12	0	8	1	0*	13	14	6	9
After 2-opt-4th	10	9	0*	6	1	2	13	0*	11	14	3	8	5	4	0*	7	12
After 2-opt-5th	8	14	9	0*	12	2	10	3	0*	7	4	11	13	1	0*	5	6
After 2-opt-6th	2	3	0*	11	1	6	14	0*	5	9	4	0*	8	13	12	7	10
After 2-opt-7th	9	6	1	2	0*	11	13	10	8	0*	14	4	7	0*	3	12	5
After 2-opt-8th	14	7	10	2	0*	8	6	12	0*	5	4	3	1	0*	11	9	13
After 2-opt-9th	3	7	13	0*	1	9	14	8	0*	11	4	10	0*	5	6	12	2
After 2-opt-10th	11	13	0*	1	14	5	12	0*	8	10	9	2	6	7	3	0*	4
After 2-opt-11th	2	13	10	0*	7	9	12	14	0*	1	6	5	11	0*	8	3	4
After 2-opt-12th	9	10	0*	1	5	11	6	4	0*	7	14	12	0*	13	8	2	3
After 2-opt-13th	10	2	4	0*	6	7	14	0*	13	1	5	8	0*	3	12	9	11
After 2-opt-14th	1	10	7	0*	2	3	9	13	14	0*	6	5	8	0*	4	12	11
After 2-opt-15th	8	6	14	0*	1	7	9	4	0*	2	11	13	0*	10	3	5	12
After 2-opt-16th	3	11	0*	10	4	7	12	0*	13	5	2	0*	14	6	8	9	1
After 2-opt-17th	4	9	14	1	0	2	10	7	0*	13	12	8	3	0*	6	5	11
After 2-opt-18th	6	10	11	0*	14	12	4	3	0*	1	2	13	7	9	0*	8	5
After 2-opt-19th	11	1	0*	10	8	9	2	13	3	0*	14	5	12	0*	4	7	6
After 2-opt-20th	9	11	7	0*	6	10	14	0*	5	1	8	3	4	13	0*	2	12

Note: Numbers against asterisk mark denote the depot and the end of one route.

TABLE 15: Updated particles.

1st	11	2	9	0*	3	5	8	12	13	0*	10	14	7	0*	1	6	4
2nd	7	8	0*	12	10	13	6	4	0*	14	3	11	9	0*	1	5	2
3rd	1	11	4	*	7	9	10	6	0*	8	13	3	0*	5	14	2	12
4th	5	14	0*	10	9	13	4	6	0*	2	3	11	0*	12	1	7	8
5th	8	12	2	3	0*	4	5	7	6	0*	10	11	13	0*	9	14	1
6th	8	7	5	0*	11	10	12	6	13	4	0*	14	3	2	0*	9	1
7th	1	7	12	0*	8	10	5	6	14	0*	3	9	2	0*	4	11	13
8th	11	8	14	9	0*	3	13	10	4	0*	5	12	1	0*	6	7	2
9th	6	13	0*	10	1	14	12	0*	4	9	11	3	0*	8	7	2	5
10th	6	4	2	0*	9	12	8	3	10	0*	14	5	11	0*	1	7	13
11th	4	2	0*	9	8	13	1	0*	3	5	12	7	0*	6	10	14	11
12th	4	5	12	2	0*	1	8	14	7	0*	6	13	10	9	0*	3	11
13th	14	2	7	0*	8	4	1	10	9	0*	5	13	6	0*	3	12	11
14th	1	2	8	9	0*	6	11	14	10	12	0*	5	7	13	0*	3	4
15th	6	14	13	0*	2	7	12	8	0*	10	9	3	1	0*	11	5	4
16th	14	10	8	0*	2	5	12	11	3	0*	7	13	4	0*	6	1	9
17th	2	13	3	5	0*	1	8	11	6	0*	12	9	0*	10	4	14	7
18th	8	13	1	2	0*	7	14	6	0*	10	11	4	12	0*	9	5	3
19th	6	5	11	0*	3	13	4	14	0*	7	10	12	0*	8	2	9	1
20th	14	1	2	0*	3	12	4	0*	13	7	8	6	5	0*	11	10	9

Note: Numbers against asterisk mark denote the depot and the end of one route.

The results are provided in Table 15. Subsequently, the algorithm calculates the fitness value, personal best, and global best of the updated particles until $|fitness(n +$

$1) - fitness(n)|/|fitness(n + 1)| \leq 0.05$ or the iterative value exceeds the value of the original iterative step.

TABLE 16: Parameters of the algorithm.

Parameters' name of Hybridization algorithm of DPSO with VNS	Value
Population of particles (N)	20
The length of per particle (L)	14
Weight Value (C)	0.6
Iteration	50
Learning Value (C_p)	1.5
Social Value (C_g)	2

TABLE 17: Dispatching plan.

Route-ID	Routes	Cost of per route
R1	9-2-1-8	5.4 km
R2	3-13-14	4.7 km
R3	10-11-12	6.2 km
R4	4-6-5-7	9.33 km
Total Cost		25.63 km

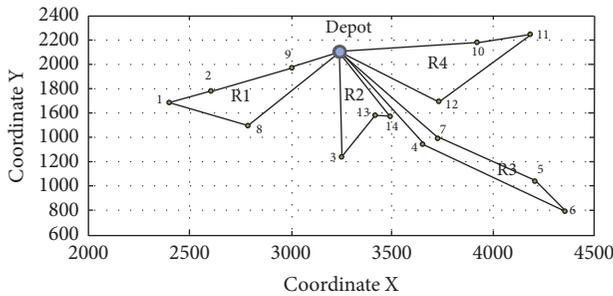


FIGURE 6: Dispatching route of the hybridization algorithm of DPSO with VNS.

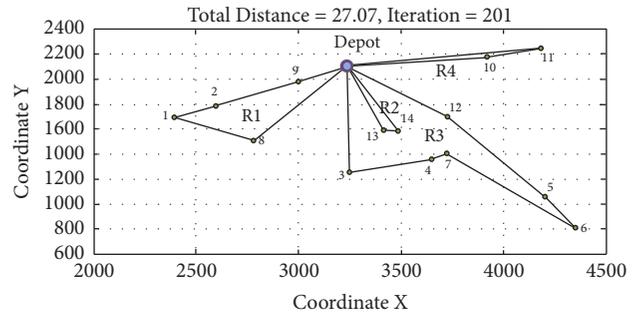


FIGURE 7: Result of DPSO algorithm.

When the conditions of convergence are met, the operation is complete. Otherwise, return to process in Section 6.1 and implement the iterative operation again.

6.6. *Computational Result and Analysis.* The parameters of the hybridization algorithm of DPSO with VNS are included in Table 16.

The dispatching plan is presented in Table 17 and the dispatching routes are shown in Figure 6.

Computation using the hybridization algorithm of DPSO with VNS produces four routes for dispatching, route 1= [9-2-1-8], route 2= [3-13-14], route 3= [10-11-12], and route 4 = [4-6-5-7]. Furthermore, the cost per route scaled by the distance corresponds to 5.4 km, 4.7 km, 6.2 km, and 9.33 km. The total cost is 25.63 km.

In order to test the performance of the hybridization algorithm of DPSO with VNS, Goksal et al. made an overall comparison with Reactive Tabu Search (RTS), Large Neighborhood Search (LNS), Particle Swarm Optimization (PSO), Ant Colony System (ACS), Parallel Iterative Local Search (PILS), and Adaptive Memory Methodology (AMM). Since they are the most effective heuristics proposed in the literature. According to the indexes of average deviation to the best known solution, average computation time, scaled

average computation time over 40 instances, and number of best known solutions found, the comprehensive performance of hybridization algorithm of DPSO with VNS is the best. For the computational results of the comparison, see Goksal et al. [21].

In this study, we implement Discrete Particle Swarm Optimization (DPSO) and Genetic algorithm (GA) separately over the case study data set to get a more comprehensive view of the performance of the hybridization algorithm of DPSO with VNS. The results are provided in Figures 7 and 8. DPSO is considered in the comparison with the purpose of presenting the performance of VNS. We can see from Figure 7 that the cost of the solution without VNS is 27.07 km. The result of the hybridization algorithm of DPSO with VNS (25.63 km) dropped by 5.3%. GA [15] is a population-based heuristic which can be used to solve the problem like SBRP. Y. Li et al. reported the performance in their study. Thus, in order to make a direct comparison, we run a GA program over the same data set of case study. From Figure 8, we can see that the total cost is 26.33 km, while the hybridization algorithm of DPSO with VNS brings a drop by 2.7%. This analysis indicates that the hybridization algorithm of DPSO with VNS outperforms the other existing heuristics in the SBRP literature.

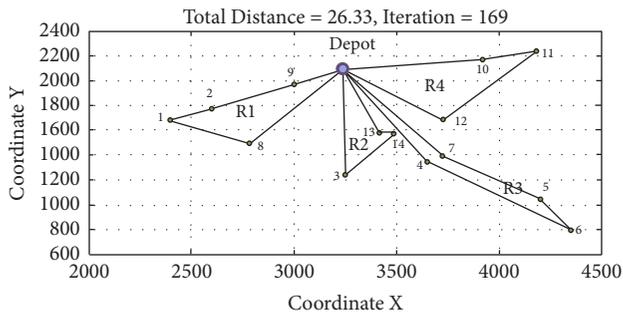


FIGURE 8: Result of GA.

7. Concluding Remarks

We quantify the station state using the bicycle net flow by calculating the difference between the inflow and the outflow. The maximum daily demand can be determined by the minimum value on the state curve. It should be noted that this minimum daily demand reveals the imbalance among the bicycle arrival and departure, rather than the real demand of the static rebalancing.

An extended VRPSPD program is developed to model the bike-sharing static rebalancing problem considering the number of bicycles that need repair. We employed a hybrid DPSO-VNS algorithm to solve the proposed model. The computational result shows that the effectiveness of the algorithm is attributed to a combination of the following reasons. First, the idea of vehicle orientation ensures that each route only covers a restricted area. Second, the quality of the solution is improved by the cheapest insertion heuristic and 2-opt method, both of which are applied during the route construction. Third, the DPSO mechanism is capable of generating diverse solutions and maintaining the best solution found during the iterative process. Moreover, this approach is flexible, because it is observed that it can be applied to the related problems by modifying the initial solution and a feasibility check of neighborhood methods. The approach can be easily extended and adjusted by adding or replacing these neighborhood methods.

Some aspects may improve the performance of the proposed algorithm, such as the parameter optimization and programming implementation. Although the PSO parameter set used in this study was obtained from a preliminary experiment, it may not be the best one. In addition, the implementation of the algorithm may be further improved. These aspects would be further investigated in future studies.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research has been substantially supported by research grants from the National Natural Science Foundation Council of China (51408058, 51338002), grants from the China Scholarship Council (201608430161), and Open Fund of Engineering Research Center of Catastrophic Prophylaxis and Treatment of Road & Traffic Safety (Changsha University of Science & Technology), Ministry of Education, under grant number kfj130301.

References

- [1] M. Dell'Amico, E. Hadjicostantinou, M. Iori, and S. Novellani, "The bike sharing rebalancing problem: Mathematical formulations and benchmark instances," *OMEGA - The International Journal of Management Science*, vol. 45, pp. 7–19, 2014.
- [2] R. Alvarez-Valdes, J. M. Belenguer, E. Benavent et al., "Optimizing the level of service quality of a bike-sharing system," *OMEGA - The International Journal of Management Science*, vol. 62, pp. 163–175, 2016.
- [3] A. A. Kadri, I. Kacem, and K. Labadi, "A branch-and-bound algorithm for solving the static rebalancing problem in bicycle-sharing systems," *Computers & Industrial Engineering*, vol. 95, pp. 41–52, 2016.
- [4] S. Salhi and G. Nagy, "A cluster insertion heuristic for single and multiple depot vehicle routing problems with backhauling," *Journal of the Operational Research Society*, vol. 50, no. 10, pp. 1034–1042, 1999.
- [5] P. Angeloudis, J. Hu, and M. G. H. Bell, "A strategic repositioning algorithm for bicycle-sharing schemes," *Transportmetrica A: Transport Science*, vol. 10, no. 8, pp. 759–774, 2014.
- [6] M. Benchimol et al., "Balancing the stations of a self service 'bike hire' system," *RAIRO - Operations Research*, vol. 45, no. 1, pp. 37–61, 2011.
- [7] J. H. Lin and T. C. Chou, "A geo-aware and VRP-based public bicycle redistribution system," *International Journal of Vehicular Technology*, vol. 2012, 2012.
- [8] T. Raviv, M. Tzur, and I. A. Forma, "Static repositioning in a bike-sharing system: models and solution approaches," *EURO Journal on Transportation and Logistics*, vol. 2, no. 3, pp. 187–229, 2013.
- [9] D. Chemla et al., "Bike sharing system: solving the static rebalancing problem," Article ID hal-00726617, 2013.
- [10] W. Y. Szeto, Y. Liu, and S. C. Ho, "Chemical reaction optimization for solving a static bike repositioning problem," *Transportation Research Part D: Transport and Environment*, vol. 47, pp. 104–135, 2016.
- [11] F. Cruz, A. Subramanian, B. P. Bruck, and M. Iori, "A heuristic algorithm for a single vehicle static bike sharing rebalancing problem," *Computers & Operations Research*, vol. 79, pp. 19–33, 2017.
- [12] S. C. Ho and W. Y. Szeto, "A hybrid large neighborhood search for the static multi-vehicle bike-repositioning problem," *Transportation Research Part B: Methodological*, vol. 95, pp. 340–363, 2017.
- [13] S. C. Ho and W. Y. Szeto, "Solving a static repositioning problem in bike-sharing systems using iterated tabu search," *Transportation Research Part E: Logistics and Transportation Review*, vol. 69, pp. 180–198, 2014.

- [14] I. A. Forma, T. Raviv, and M. Tzur, "A 3-step math heuristic for the static repositioning problem in bike-sharing systems," *Transportation Research Part B: Methodological*, vol. 71, pp. 230–247, 2015.
- [15] Y. Li, W. Y. Szeto, J. Long, and C. S. Shui, "A multiple type bike repositioning problem," *Transportation Research Part B: Methodological*, vol. 90, pp. 263–278, 2016.
- [16] J. Schuijbroek, R. C. Hampshire, and W.-J. van Hoesve, "Inventory rebalancing and vehicle routing in bike sharing systems," *European Journal of Operational Research*, vol. 257, no. 3, pp. 992–1004, 2017.
- [17] W. Y. Szeto and C. S. Shui, "Exact loading and unloading strategies for the static multi-vehicle bike repositioning problem," *Transportation Research Part B: Methodological*, vol. 109, pp. 176–211, 2018.
- [18] T. Bulhões, A. Subramanian, G. Erdoğan, and G. Laporte, "The static bike relocation problem with multiple vehicles and visits," *European Journal of Operational Research*, vol. 264, no. 2, pp. 508–523, 2018.
- [19] J. Dethloff, "Vehicle routing and reverse logistics: the vehicle routing problem with simultaneous delivery and pick-up," *OR Spectrum*, vol. 23, no. 1, pp. 79–96, 2001.
- [20] H. Min, "The multiple vehicle routing problem with simultaneous delivery and pick-up points," *Transportation Research Part A: General*, vol. 23, no. 5, pp. 377–386, 1989.
- [21] F. P. Goksal, I. Karaoglan, and F. Altıparmak, "A hybrid discrete particle swarm optimization for vehicle routing problem with simultaneous pickup and delivery," *Computers & Industrial Engineering*, vol. 65, no. 1, pp. 39–53, 2013.
- [22] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system," *Pervasive and Mobile Computing*, vol. 6, no. 4, pp. 455–466, 2010.
- [23] Y. Feng and S. Wang, "A Forecast for Bicycle Rental Demand Based on Random Forests and Multiple Linear Regression," pp. 2–6, 2017.
- [24] S. V. Kumar, K. C. Dogiparthi, L. Vanajakshi, and S. C. Subramanian, "Integration of exponential smoothing with state space formulation for bus travel time and arrival time prediction," *Transport*, vol. 4142, pp. 1–10, 2015.
- [25] F. Lin, L. Zhang, H. Xu, and Y. Sun, "An Improved Exponential Smoothing Model on Rental Trends Prediction of Public Bicycle Stations," in *Proceedings of the 2016 IEEE 14th IEEE International Conference on Dependable, Autonomic and Secure Computing. DASC 2016, 2016 IEEE 14th International Conference Pervasive Intelligence and Computer. PICom 2016, 2016 IEEE 2nd International Conference Big Data*, pp. 437–441, 2016.
- [26] H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications," *Environmental Modelling & Software*, vol. 15, no. 1, pp. 101–124, 2000.
- [27] C. Liu, S. C. H. Hoi, P. Zhao, and J. Sun, "Online ARIMA algorithms for time series prediction," in *AAAI'16 Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (IJCAI)*, pp. 1867–1873, 2016.
- [28] D. Ö. Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, pp. 586–594, 2010.
- [29] G. Berbeglia, J.-F. Cordeau, I. Gribkovskaia, and G. Laporte, "Rejoinder on: Static pickup and delivery problems: a classification scheme and survey," *TOP*, vol. 15, no. 1, pp. 45–47, 2007.
- [30] S. N. Parragh, K. F. Doerner, and R. F. Hartl, "A survey on pickup and delivery problems. Part I: Transportation between customers and depot," *Journal für Betriebswirtschaft*, vol. 58, no. 2, pp. 81–117, 2008.
- [31] T. J. Ai and V. Kachitvichyanukul, "A particle swarm optimization for the vehicle routing problem with simultaneous pickup and delivery," *Computers & Operations Research*, vol. 36, no. 5, pp. 1693–1702, 2009.
- [32] Y. Gajpal and P. Abad, "An ant colony system (ACS) for vehicle routing problem with simultaneous delivery and pickup," *Computers & Operations Research*, vol. 36, no. 12, pp. 3215–3223, 2009.
- [33] M. Reimann, K. Doerner, and R. F. Hartl, "D-ants: Savings based ants divide and conquer the vehicle routing problem," *Computers & Operations Research*, vol. 31, no. 4, pp. 563–591, 2004.
- [34] J. Li, P. M. Pardalos, H. Sun, J. Pei, and Y. Zhang, "Iterated local search embedded adaptive neighborhood selection approach for the multi-depot vehicle routing problem with simultaneous deliveries and pickups," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3551–3561, 2015.
- [35] L. Lan and W. Xia, "Application Based on PSO of Hopfield Neural Network Algorithm in Dynamic VRPSPD," in *Proceedings of the 12th IEEE International Conference on E-Business Engineering, ICEBE 2015*, pp. 107–113, China, October 2015.

Research Article

Impact of a New Metro Line: Analysis of Metro Passenger Flow and Travel Time Based on Smart Card Data

Xiao Fu ¹ and Yu Gu ²

¹School of Transportation, Southeast University, Nanjing 210096, China

²Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Centre of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing 210096, China

Correspondence should be addressed to Xiao Fu; fuxiao@seu.edu.cn

Received 27 May 2018; Accepted 6 August 2018; Published 19 August 2018

Academic Editor: David Z. W. Wang

Copyright © 2018 Xiao Fu and Yu Gu. 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.

Over the past few decades, massive volumes of smart card data from metro systems have been used to investigate passengers' mobility patterns and assess the performance of metro network. With the rapid development of urban rail transit in densely populated areas, new metro lines are constantly designed and operated in recent years. The appearance of new metro lines may significantly affect passenger flow and travel time in the metro network. In this study, smart card data of metro system from Nanjing, China, are used to study the changes of metro passenger flow and travel time due to the operation of a new metro line (i.e., Line 4, opened on 18 January 2017). The impact of the new metro line on passenger flow distribution and travel time in the metro network is first analysed. As commuting is one of the major purposes of metro trips, the impact of the new metro line on commuters' trips is then explicitly investigated. The results show that the new metro line influences passenger flow, travel time, and travel time reliability in the metro network and has different impacts on different categories of commuters.

1. Introduction

Multimodal transit networks provide convenient and efficient ways for individuals' daily travel in congested urban areas. Metro system, as a major component of multimodal transit system, has been significantly developed in recent years in many cities with large populations. Taking Nanjing, one of the most densely populated cities in China, for instance, 6 new metro lines have been constructed and operated in Nanjing in the past decade and 11 new metro lines are planned to be developed in the next five years.

To study the passenger flow characteristics and evaluate the performance of urban transit network, a substantial body of literature has been developed on analysing the distribution of passenger flow and travel time based on smart card data. However, limited effort has been devoted to investigating the impact of new metro line on metro network and passengers. The development of new metro lines changes the topology of metro network and influences the accessibility of city area. The operation of new metro lines may affect passenger flows of existing metro stations. Passengers' travel times and travel

time reliability may also change as a result of the expansion of metro network. As commuters are the majority of metro users in peak hours and are expected to benefit from the opening of new lines, the impact of new metro line on commuters' trips should be explicitly studied. In the literature, little attention has been given to the changes of commuters' spatial/temporal distributions and travel times after the opening of new metro lines.

In this study, a comprehensive analysis of metro passenger flow and travel time is presented on the basis of smart card data from Nanjing, China. The impact of new metro line on metro passenger flow and travel time is first analysed on the level of whole metro network. Spatial and temporal distributions of metro passenger flow in peak and nonpeak hours are investigated before and after the opening of the new line. The changes of passenger travel time and travel time reliability are also studied. To investigate the impact of the new metro line on commuters' trips, this study identifies and classifies the passengers commuting by metro and investigates the impact of the new metro line on flow distribution, travel time,

and travel time reliability of different categories of metro commuters.

The outline of this paper is as follows. Section 2 presents a review of studies using smart card data and studies discussing the impact of new rail transit lines. Data used in this study are described in Section 3. Impacts of new metro line on passenger flow and passenger travel time are investigated in Sections 4 and 5. Conclusions and recommendations for future research are given in the last section.

2. Literature Review

2.1. Applications of Smart Card Data. Smart card data, which are passively collected by automated fare collection systems with the use of smart cards for transit services, are important data source for transport researchers. Because of the large quantity and high precision of smart card data, many researchers believe that such innovative data source can complement or even replace the traditional collected transport data such as travel survey data [1, 2]. Existing transport studies based on smart card data mainly focus on three areas: analysis of passenger flow patterns, investigation of travel behaviours, and assessment of transit network performance.

Many studies are devoted to investigating the characteristics of metro passenger flow patterns based on smart card data. Smart card data record the entrance/exit stations and entrance/exit times of metro passengers, so the data are widely used to investigate and visualize the spatial and temporal distributions of metro passenger flow [3–5]. Based on the historical entrance and exit passenger flow records, the future passenger flow can also be predicted with smart card data [6, 7].

Smart card data can also be used to analyse travel behaviours and extract travel patterns of transit users. Travel behaviours in transit networks can be investigated such as transit ridership and origin-destination (OD) choices [8, 9], train and route choices [10, 11], and transfer behaviours [12]. Based on the investigation of travel behaviours, many studies are conducted to extract individuals' spatial/temporal travel patterns [13–15] and mobility patterns [16–18]. The dynamics of travel behaviours and the variation of travel pattern were also studied [19–22]. Some other studies focused on the regularity of individual travel pattern, using smart card data to classify transit passengers [23, 24].

In addition, smart card data can also be applied to investigate transit network performance such as the passenger travel time and the service reliability of transit systems. For instance, some studies investigated the spatial and temporal distribution of trip travel times [25], travel time differences between different transit modes [12], and splitting and estimation of metro travel time [26]. The service reliability of transit systems can also be derived based on smart card data. Comparing passengers' entrance/exit times to the transit schedules, researchers can select schedule deviation of vehicles or at-stop waiting time of passengers to measure the service reliability [27]. The concept of travel time reliability is widely employed to evaluate service reliability of transit systems from the perspective of individual passengers. The

travel time reliability metrics include the schedule-based metrics such as excess journey time and reliability buffer time [28, 29], as well as statistical metrics such as standard deviation (SD) and coefficient of variation (CV) of travel time [4, 30–33].

2.2. Impact of New Rail Transit Lines. Construction of a new rail transit line may result in changes of individuals' travel choice behaviour. The change in mode choice attracts much attention in the literature. Senior [34] analysed the modal shift from bus and private vehicles to the light rail transit based on household survey and census data. Cao and Schonher [35] explored the impact of new light rail transit line on transit use. Harvey et al. [36] investigated people's willingness to travel by the newly built high speed rail line. Gadziński and Radzinski [37] combined household survey data with property sales transactions data to investigate the effect of a new tram on households' mode choice behaviour. In addition to studying the mode choice change, Saidi, Wirasinghe, and Kattan [38] assessed the changes in passengers' cost after introducing a new rail line based on job distribution data and origin-destination data. Hong, Boarnet, and Houston [39] adopted sociodemographic data, travel survey data, and activity data to investigate the impact of a new light rail transit line on individuals' travel behaviour.

Besides the studies focusing on the change of travel behaviour, a few studies reveal the impact of new metro lines on passenger flow. Li [40] analysed changes in passenger flow pattern caused by different types of new metro lines. Based on smart card data, Liu, Yao, and Li [41] analysed the relationship between metro network expansion, land use of station neighbourhood, and the growth of station passenger flow.

However, existing studies mainly focus on the change of passenger flow distribution on the network level, while little attention has been paid to the impact of new lines on different individuals. Transportation analyses which are not conducted from individuals' perspective may be biased as they fail to depict individuals' real travel experience [42]. Some studies have been made to investigate the performance of urban transportation network from the perspective of individual travellers using different kinds of data sources [33, 43]. In densely populated urban areas, as a large proportion of passengers' daily travels are made by metro, the opening of a new metro line may significantly influence some individuals' daily travel patterns while other individuals may not be affected. For giving a comprehensive analysis of the impact of new metro line on metro passenger flow and travel time, the analysis in this study is first conducted on the network level to present an overview of the metro system and investigate the changes of passenger flow and travel time after the new line opening. Further analysis focuses on metro commuters. Effort has been devoted to categorizing commuters based on their OD choices and evaluating the impact of the new metro line on different categories of commuters.

3. Study Area and Data Collection

Nanjing is one of the largest cities in China with a population of about 8 million. In 2005, the first metro line

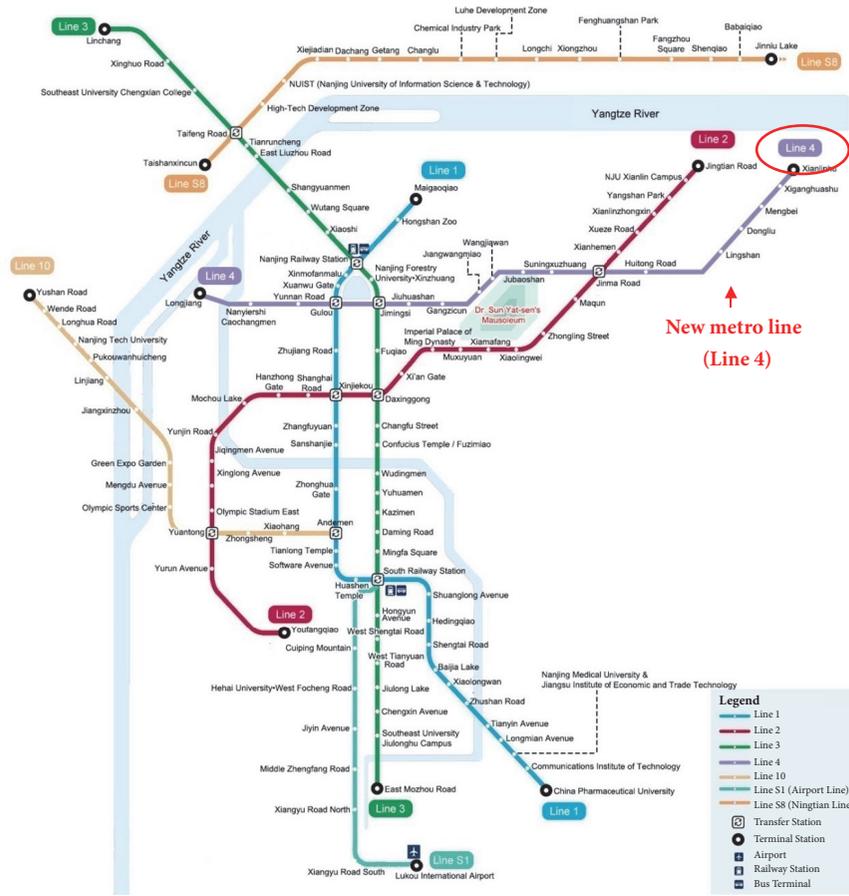


FIGURE 1: Nanjing metro network.

TABLE 1: Examples of smart card data used in this study.

Time of Entry	Station of Entrance	Time of Exit	Station of Exit	Card ID
2017/1/1 10:10:21	14	2017/1/1 10:30:21	7	976072959331
2017/1/1 20:27:50	14	2017/1/1 20:34:52	16	976072959320
2017/1/1 11:56:56	37	2017/1/1 12:46:34	101	993172764630
2017/1/1 21:10:28	14	2017/1/1 22:03:25	50	970672806438

began operation in Nanjing. In the past decade, Nanjing metro network expands rapidly. An overview of Nanjing metro network is extracted from Internet and shown in Figure 1. Line 4 of Nanjing metro system was opened on 18 January 2017, which is the latest line operated. After the opening of Line 4, there are 7 metro lines and 139 stations in operation in Nanjing. The operating mileage of Nanjing metro network is 258 km, and the average daily ridership in January 2017 is about 2.1 million.

To explore the impact of new metro line (i.e., Line 4) on metro passenger flow and travel time in densely populated urban areas, the data used in this study are records of smart cards used by metro passengers in Nanjing from January to February 2017, 61 million records in total. Data of the two months are used because the two months cover the periods before and after the opening of Line 4. Data of two

consecutive months guarantee the stability of passengers' travel choice behaviour.

Smart card data contain complete entrance and exit records (including location and time) of each passenger, as shown in Table 1. Take the first record in Table 1 as an example; the record indicates that the passenger entered the metro system from station No. 14 (i.e., Nanjing Railway Station) at 10:10:21 on 1 January 2017 and exited from station No. 7 (i.e., Sanshanjie) at 10:30:21. Data cleaning is conducted to exclude incomplete and invalid data before analysis.

4. The Impact of New Metro Line on the Metro Network

The operation of the new metro line may influence metro passenger flow distribution. It can be seen from Figure 1 that Line 4 has intersections with three lines (i.e., Lines 1, 2,

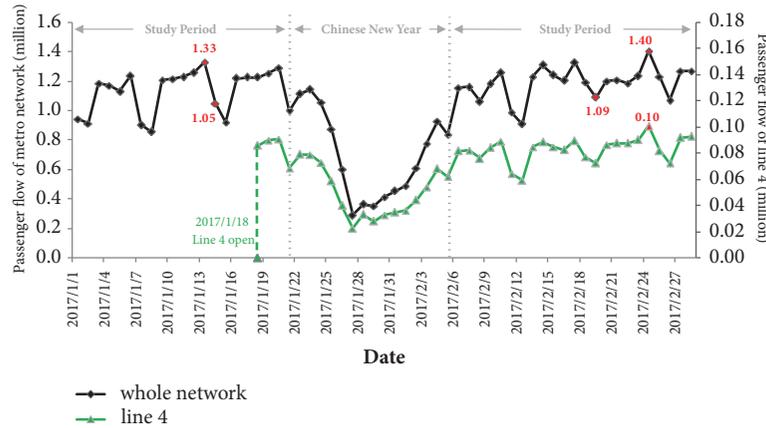


FIGURE 2: Daily metro passenger flow before and after Line 4 opening.

and 3) which pass through central areas of the city. Before the opening of Line 4, the three lines took the majority of passenger flows in the metro network. Line 4 is expected to share passenger flows with Lines 1-3 and alleviate the crowding effect at some metro stations in central business district (CBD). Thus, a need is evident to investigate the change of passenger flow in metro network after Line 4 opening.

In this section, spatial and temporal analysis of passenger flow are conducted to show the passenger flow distributions in the whole metro network before and after the opening of Line 4. Travel time and travel time reliability in the metro network are also investigated to show the impact of the new line.

4.1. Metro Passenger Flow Distribution

4.1.1. Temporal Distribution of Metro Passenger Flow. On the basis of smart card data from 1 January to 28 February 2017, the daily flows of passengers using smart cards on Line 4 and on the whole metro network are shown in Figure 2. It is obvious from Figure 2 that the passenger flow varies from day to day. The passenger flow on workday is significantly higher than that at weekend, which implies that a large portion of metro passengers are commuters. Thus, in Section 5, explicit analyses of commuters' flow distributions and travel times are given.

It can also be seen from Figure 2 that the number of passengers increases after Line 4 opening. Before the opening of Line 4, the maximum daily passenger flow of the whole metro network is 1.33 million on workday and 1.05 million at weekend. After the opening of Line 4, the maximum passenger flows on workday and at weekend increase to 1.40 million and 1.09 million, respectively. The maximum passenger flow of Line 4 reaches 0.10 million on 24 February 2017.

Please note that passenger flow decreases dramatically during 21 January and 5 February as many residents leave Nanjing during Chinese New Year. Thus, data of the period affected by Chinese New Year are excluded from analysis

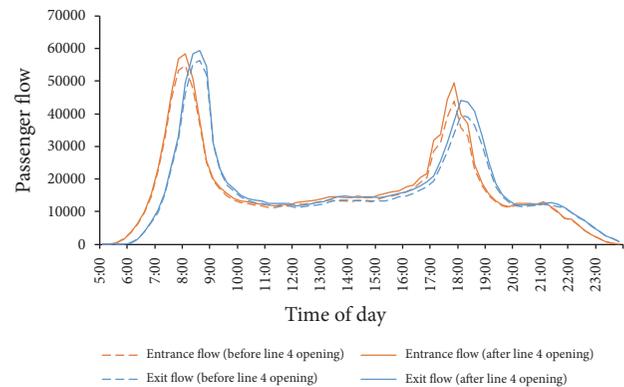


FIGURE 3: 15 min metro passenger flow distribution before and after Line 4 opening.

in this study to facilitate essential ideas without loss of generality. The study period is then divided into two parts: (a) 1 Jan 2017 to 17 Jan 2017 (the period before Line 4 opening); (b) 18 Jan 2017 to 20 Jan 2017 and 6 Feb 2017 to 28 Feb 2017 (the period after Line 4 opening).

In addition to the daily fluctuation of passenger flow, the entrance/exit flow in metro network also varies greatly at different times of day. Figure 3 shows the average 15 min passenger flow on workdays before and after Line 4 opening. Increase of entrance/exit flow can be found after Line 4 opening due to the expansion of network capacity. It can also be seen from Figure 3 that there are two peaks of passenger flow on workdays, i.e., 7:00-9:00 in the morning and 17:00-19:00 in the afternoon. To capture the different conditions of metro network, it is necessary to analyse metro passenger flow and travel time in both peak hours and nonpeak hours.

4.1.2. Spatial Distribution of Metro Passenger Flow. Figure 4 shows the average daily exit flows of metro stations before and after the opening of Line 4 based on kernel density estimate.

The passenger flows at the stations highlighted in red circles vary significantly due to the opening of Line 4. It can be seen that the crowding at the stations in CBD (i.e., the red

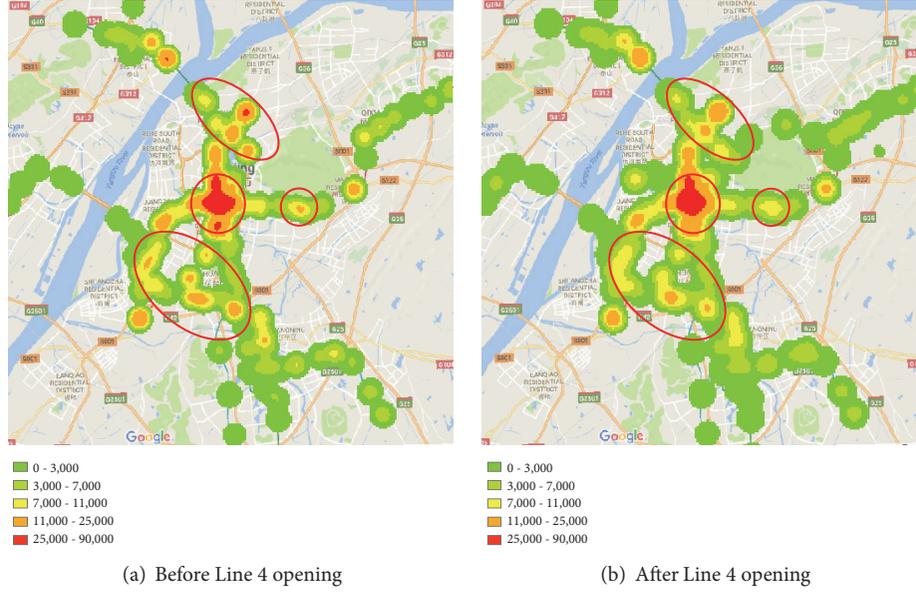


FIGURE 4: Spatial distribution of metro passenger flow before and after Line 4 opening (map source: Google Maps).

circle in the centre of the figure) is alleviated after the opening of Line 4. Several stations on other lines (highlighted in other red circles) share fewer passenger flows after the opening of Line 4. Thus, Line 4 attracts more individuals to metro network while contributing to the whole network by sharing a large number of passengers and reducing the crowding at several existing metro stations.

To explicitly investigate the impact of Line 4 on spatial distribution of passenger flow, the metro stations significantly affected by Line 4 are identified as follows.

Denote the mean and the SD of daily passenger flow from station i to station j before Line 4 opening as f_{ij}^b and σ_{ij}^b , respectively:

$$f_{ij}^b = \frac{\sum_{n=1}^{N^b} f_{ij}^{nb}}{N^b}, \quad (1)$$

$$\sigma_{ij}^b = \sqrt{\frac{\sum_{n=1}^{N^b} (f_{ij}^{nb} - f_{ij}^b)^2}{(N^b - 1)}}, \quad (2)$$

where N^b is the number of days of the study period before Line 4 opening, and f_{ij}^{nb} is the passenger flow on day n .

The mean of passenger flow from station i to station j after Line 4 opening (denoted as f_{ij}^a) can be obtained as

$$f_{ij}^a = \frac{\sum_{n=1}^{N^a} f_{ij}^{na}}{N^a}, \quad (3)$$

where N^a is the number of days of the study period after Line 4 opening.

In this study, if f_{ij}^a exceeds the normal fluctuation ranges

$$\begin{aligned} f_{ij}^a &> f_{ij}^b + \beta \cdot \sigma_{ij}^b \\ \text{or } f_{ij}^a &< f_{ij}^b - \beta \cdot \sigma_{ij}^b, \end{aligned} \quad (4)$$

the OD stations are viewed to be affected by Line 4. β is a parameter associated with the normal fluctuation ranges. In this study, the fluctuation range is set as 95%; thus β is 1.64. Setting the destination j as Gulou (a station located in CBD with a large number of passenger flows), Figure 5 shows the stations affected by Line 4 in peak and nonpeak hours.

As shown in Figure 5, the passenger flow travelling to CBD is significantly influenced by the operation of Line 4. From Figure 5(a), it can be seen that stations on Line 3 significantly attract more passengers to CBD, while the passenger flows of many stations on Line 1 and Line 2 decreased. The reason is that Line 2 is spatially parallel to the new metro Line 4, and passengers originally travelling by Line 2 might change to use Line 4. The shift of passenger flow from Line 1 to Line 3 can be explained by the decreased travel time from Line 3 to CBD, which will be demonstrated in Section 4.2. Comparing Figures 5(a) and 5(b), it can be found that the passenger flow distribution in nonpeak hours is relatively more stable than that in peak hours, indicating that commuting trips are significantly influenced by the new metro line.

4.2. Travel Time in the Metro Network. The introduction of a new metro line probably has impacts, as expected, on passengers' activity and travel choice behaviour as the accessibility/connectivity of activity locations and transit stations changes [44]. As a reliable transport mode which does not suffer from road traffic congestion, metro is one of the best choices for commuters who work in CBD. In a city with similar size of Nanjing, the 30-minute coverage area of metro stations in CBD is important to commuters. Figure 6 shows the travel times from different metro stations to CBD with a travel time less than 30 minutes. Apparently, the number of stations within the 30-minute coverage area greatly increases due to the opening of Line 4. The travel

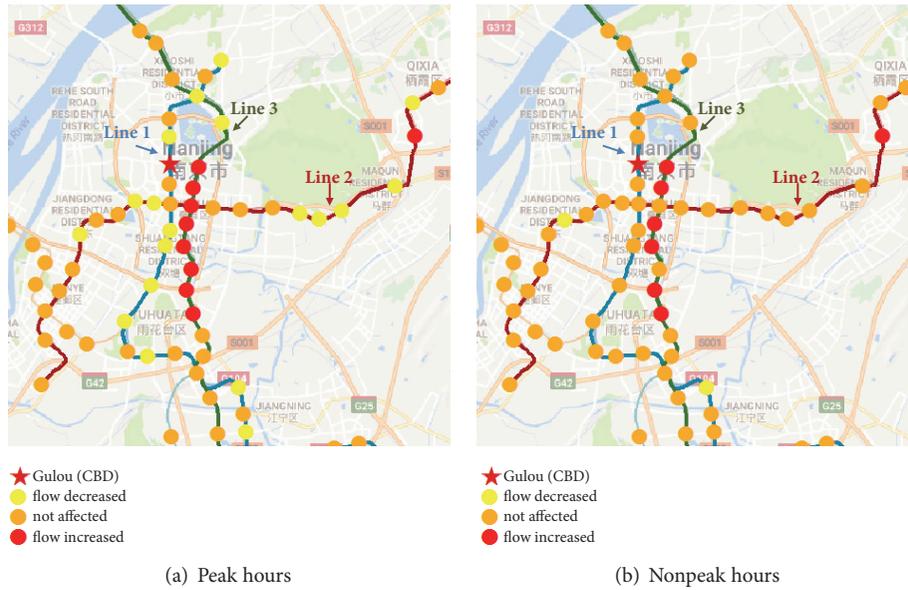


FIGURE 5: Change of passenger flow at metro stations after Line 4 opening (map source: Google Maps).

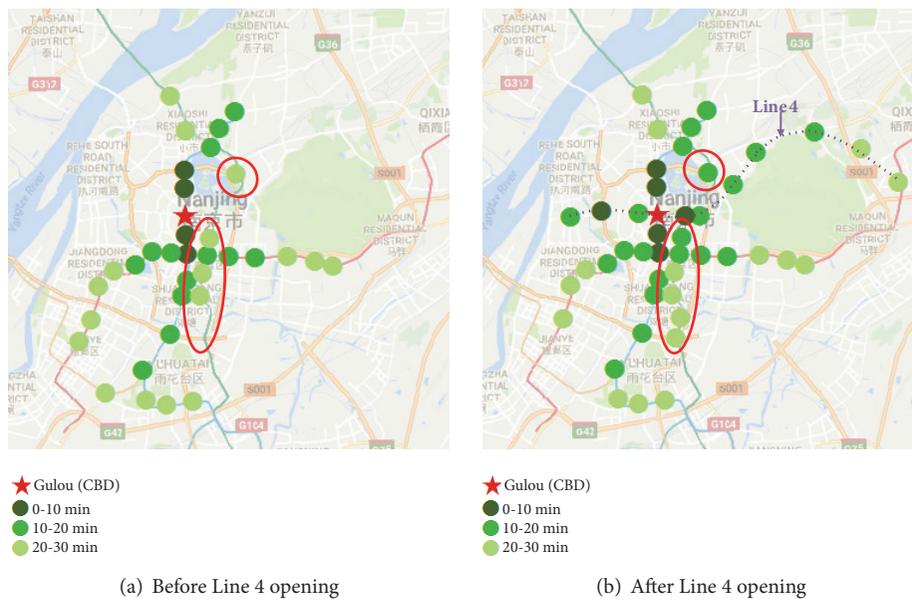


FIGURE 6: Travel times from metro stations to CBD (map source: Google Maps).

times from many stations on Line 3 (which are highlighted by red circles) to CBD also decrease after Line 4 opening, while the travel times from stations on Line 1 to CBD remain stable. Hence, many passengers choose Line 3 to travel to CBD instead of Line 1. The shift of passenger flow from Line 1 to Line 3 can be found in Figure 5.

In metropolitan areas, travel times in transit networks generally vary from day to day. Many empirical studies revealed that individuals, especially commuters, indeed consider the travel time uncertainty as a risk for their travels. To reduce the risk of late arrival, individuals may have concerns on the probability that a trip can be successfully fulfilled within a given travel time, referred to as travel time reliability

in the literature. Therefore, the travel time reliability in metro network should be explicitly explored with real data.

In this study, the coefficient of variation (CV) of travel time is selected as the measure of travel time reliability. The travel time reliability within OD pair ij , cv_{ij} , can be expressed as

$$cv_{ij} = \frac{\sigma_{ij}}{t_{ij}}, \quad (5)$$

where σ_{ij} and t_{ij} , respectively, denote the SD and the mean of travel time between i and j . A higher cv_{ij} indicates lower travel time reliability.

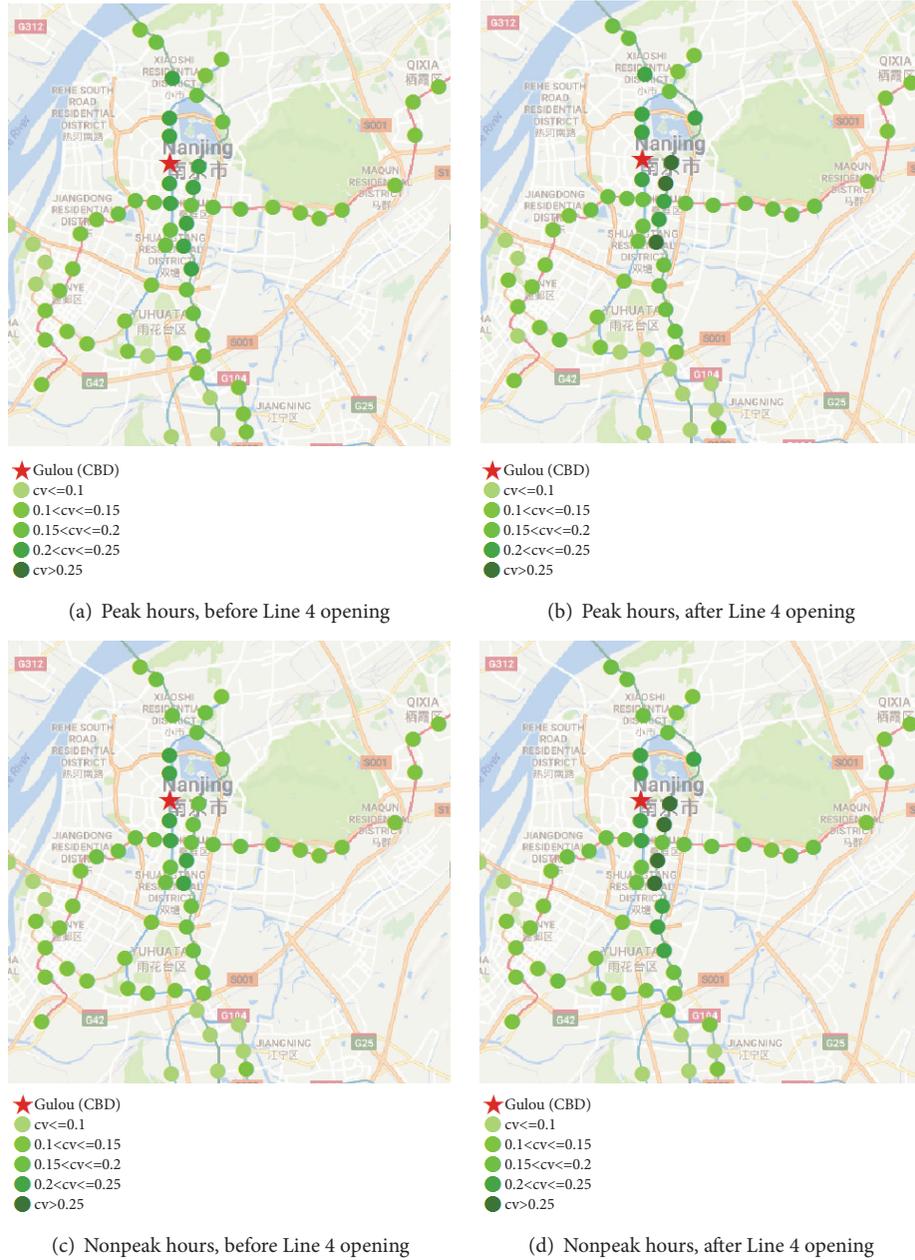


FIGURE 7: Travel time reliability from metro stations to CBD (map source: Google Maps).

Based on the trip travel times extracted from smart card data before and after the opening of Line 4, the travel time reliability from all stations to CBD can be analysed. Figure 7 shows the CVs of travel times during peak and nonpeak hours before and after Line 4 opening. A light colour of a station is with a small value of CV, which indicates the travel time from this station to CBD is quite reliable. The crowding effect during peak hours may reduce travel time reliability because some individuals are not able to get on the first arriving vehicles; however, the frequency of metro service in peak hours is higher than that in nonpeak hours. Thus, comparing Figures 7(a) and 7(c) or comparing Figures 7(b) and 7(d), it can be found that there is no identical trend for all

stations from peak to nonpeak hours in terms of travel time reliability.

Comparing Figures 7(a) and 7(b) or comparing Figures 7(c) and 7(d), it can be found that the opening of Line 4 surprisingly leads to decreased travel time reliability for many stations, especially the stations on Line 3. The finding can be attributed to the increased transfer choices introduced by Line 4. Many individuals prefer to transfer more times for shortening the travel time after Line 4 opening, while transfers bring random factors such as stochastic waiting and walking time.

The analysis in this section aggregates passengers in the whole metro network, which is unable to reveal the flow

distribution and the travel time of different passengers before and after the opening of the new line. In Section 5, analyses of passenger flow and travel time focusing on commuters with different OD choices are conducted.

5. The Impact of New Metro Line on Metro Commuters

The analysis of passenger flow on the whole metro network level indicates the large passenger flow during peak hours in workdays, implying that commuters are the majority of metro users. Therefore, the spatial and temporal distribution of metro commuter flows and commuters' travel times deserve further analysis. The travel time reliability, which is a major concern for commuters' daily travel, should also be investigated before and after the opening of the new metro line.

In this section, analysis is conducted to investigate passenger flow and travel time changes for different categories of commuters.

5.1. Classification of Commuters. As commuting is a regular activity conducted by commuters in workdays, this study defines metro commuting as metro trips conducted by passengers between fixed ODs during peak hours of workdays. The metro commuters are defined as the passengers who conduct no less than one metro commuting trip every workday.

As the commuters' choices of entrance/exit station may change after introducing Line 4, the commuters can be further divided into three groups based on the changes of their entrance/exit stations (i.e., ODs). The three groups are as follows: (1) commuters whose entrance and exit stations are not changed and not on Line 4 throughout the whole study time period; (2) commuters whose entrance and exit stations are not changed but the entrance/exit stations are intersections of Line 4 and other lines; (3) commuters whose entrance/exit stations are changed from stations on other lines to stations on Line 4 after Line 4 opening. The three groups, respectively, represent the commuters whose ODs are not affected, indirectly affected, and directly affected by the opening of Line 4.

Based on above criteria, 8543, 588, and 639 passengers are identified as groups 1, 2, and 3 commuters, respectively. The distributions and travel times of the commuters in different groups are further investigated. The impacts of new metro line on different groups are discussed.

5.2. Passenger Flow Distribution of Different Groups of Commuters

5.2.1. Temporal Distribution of Different Groups of Commuters. The temporal distributions of the three groups of commuters over a day are presented in Figure 8. From the entrance/exit passenger flow profiles, it can be found that the operation of Line 4 exerts different degrees of impacts on the three groups of commuters. A significant growth in the flow of group 3 commuters can be found after Line 4 opening. However, the

distributions of group 1 and group 2 passenger flows remain almost unchanged. The new metro line generates more commuting trips by providing more choices of O/D stations overall, but the commuters who have fixed OD stations and do not need to use the new line are hardly influenced.

As regards passenger travel time, hardly any change is observed in group 1 after Line 4 opening. A slight difference can be found between the travel times of group 2 before and after Line 4 opening. Although group 2 does not change ODs after Line 4 opening, the travel time of group 2 changes a little because their O/Ds are transfer stations with Line 4. The passenger flows in these transfer stations change after Line 4 opening which may lead to the travel time change. As for group 3, it can be seen that the travel time varies a lot after Line 4 opening. The travel times at different times of day are significantly different from the travel times before Line 4 opening. As group 3 changes O/Ds after Line 4 opening, the travel time changes accordingly.

5.2.2. Spatial Distribution of Different Groups of Commuters. Figure 9 depicts the cumulative number of trips with respect to the proportion of ordered OD pairs. It can be seen that a great number of trips are conducted between a small proportion of OD pairs. For instance, about 70% of the trips of group 1 are conducted between 10% OD pairs, implying that the majority of commuters' trips are between a few ODs. In addition, comparing to the trips in nonpeak hours, trips in peak hours are generated between fewer OD pairs, indicating that the spatial distribution of commuter flow concentrates in a limited number of ODs in peak hours. This phenomenon is more obvious in the morning peak than the afternoon peak, as commuters have more destination choices after work in the afternoon.

Comparing the distribution curves before and after Line 4 opening, we can find that the opening of Line 4 has different degrees of impact on the spatial distributions of different groups of commuters. The spatial distribution of group 1 is scarcely influenced. The spatial distribution of group 2 is less influenced in peak hours but more influenced in nonpeak hours, indicating that group 2 commuters have more OD choices in their spare time after Line 4 opening. Apparently, the impacts of Line 4 on the spatial distribution of group 3 are significant in both peak and nonpeak hours as group 3 changes O/D stations after Line 4 opening.

5.3. Travel Time Reliability of Different Groups of Commuters. It is widely recognized that the analysis of travel time reliability should focus on individual passengers [45, 46]. In this section, the travel time reliability of each commuter group is investigated. The impacts of new metro line on travel time reliability of different groups of commuters are discussed.

Being consistent with Section 4.2, the average CV of each group of commuter's travel time is obtained to indicate the metro travel time reliability. CV of individual k , cv_k , can be expressed as

$$cv_k = \frac{\sigma_k}{t_k}, \quad (6)$$

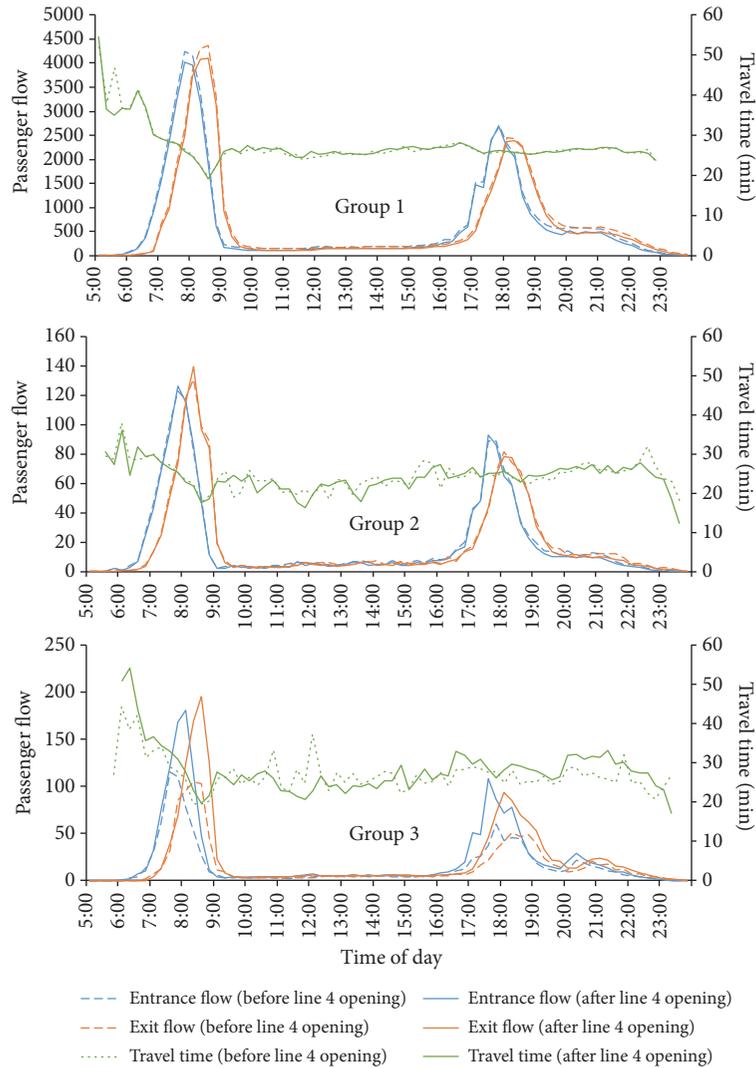


FIGURE 8: 15 min flow distributions of different groups of commuters and their travel times before and after Line 4 opening.

where σ_k and t_k , respectively, denote the SD and the mean of k 's metro travel time between his commuting OD.

The mean and the SD of commuting time of each group before and after Line 4 opening are investigated and shown in Figure 10. As linear lines fit the trends well and the slope of line can intuitively depict the travel time reliability, linear trend lines are used in Figure 10. It can be found from Figures 10(a) and 10(b) that the mean and the SD of travel time of group 1 are hardly influenced by the opening of Line 4, as the OD choices of group 1 are not affected by Line 4.

Comparing Figures 10(c) and 10(d), it can be seen that the distribution of data points in Figure 10(d) is more concentrated around the trend line, indicating that the travel time variability of group 2 is more predictable. This is because Line 4 shares a considerable number of passengers travelling through the transfer stations (i.e., group 2's O/Ds). Thus, the variability of travel time is mainly associated with travel distance instead of other incidents such as overcrowdings at stations.

From Figures 10(e) and 10(f), it can be found that the slope of trend line after Line 4 opening is obviously lower than that before Line 4 opening, indicating that the introduction of Line 4 raises the travel time reliability of group 3 commuters. Group 3 shifts to use Line 4 instead of other lines. The passenger flow on Line 4 is relatively low compared to other lines as Line 4 is newly put into operation, which guarantees high travel time reliability.

The mean and the average CV of individual travel times in each commuter group before and after Line 4 opening are listed in Table 2. It can be found that after Line 4 opening the mean travel time of group 1 and group 2 hardly changes. It indicates that for passengers who do not change OD choices the effect of introducing new metro line on average travel time is little. Note that the mean travel time of group 2 commuters slightly decreases. This is because the operation of Line 4 results in reduced number of transfers for some group 2 commuters. However, the average travel time of group 3 significantly increases after Line 4 opening. It indicates that

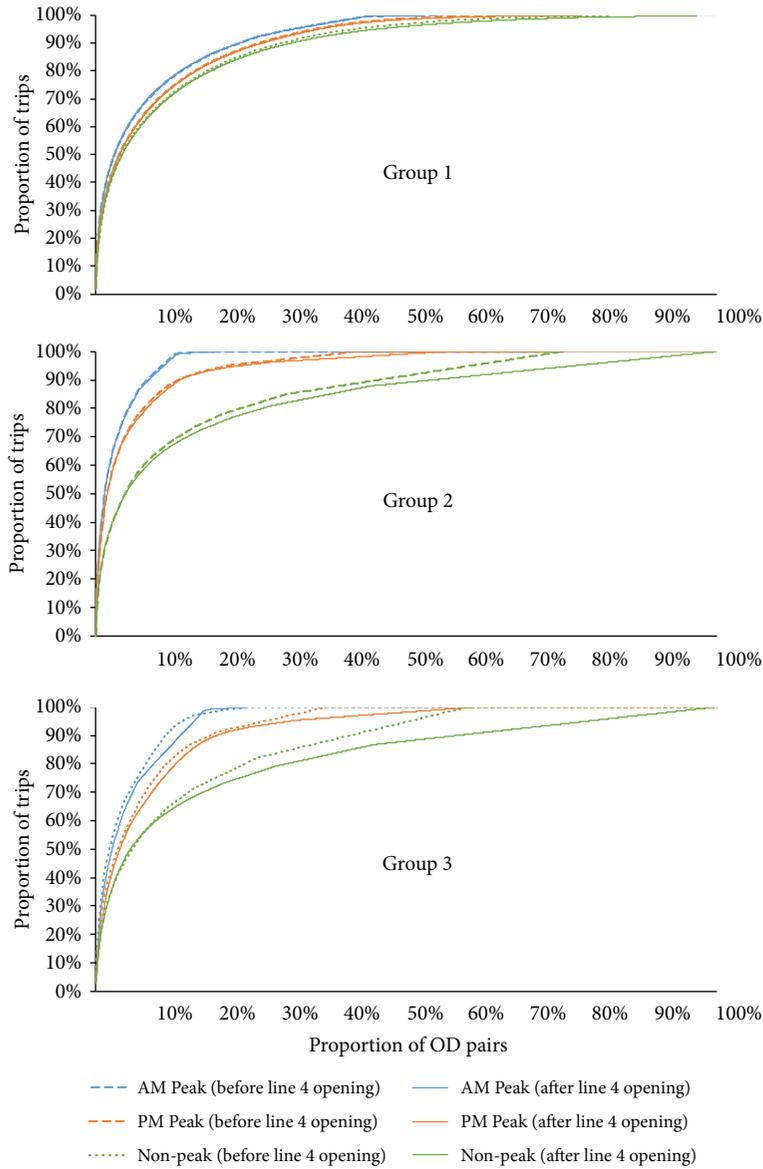


FIGURE 9: Cumulative number of trips with regard to proportion of ODs before and after Line 4 opening.

TABLE 2: Mean and average CV of commuting travel time.

	Before Line 4 opening		After Line 4 opening	
	Mean (min)	Average CV	Mean (min)	Average CV
Group 1	25.545	0.100	25.624	0.100
Group 2	25.630	0.084	25.063	0.082
Group 3	25.975	0.096	34.306	0.086

the new line makes group 3 commuters prefer travelling farther by metro, as the new stations provide commuters with more entrance/exit station choices.

It can be concluded from Table 2 that the new metro line has different levels of impacts on the travel time reliability of commuters in different groups. Introducing Line 4 hardly influences travel time reliability of commuters who do not

change their OD choices. The average CV of group 1 travel time remains stable, while the average CV of group 2 travel time slightly decreases from 0.084 to 0.082. The impact of Line 4 on group 3 travel time reliability is significant. The average CV of group 3 travel time (i.e., the travel time of commuters who shift to use new stations on Line 4) drops from 0.096 to 0.086.

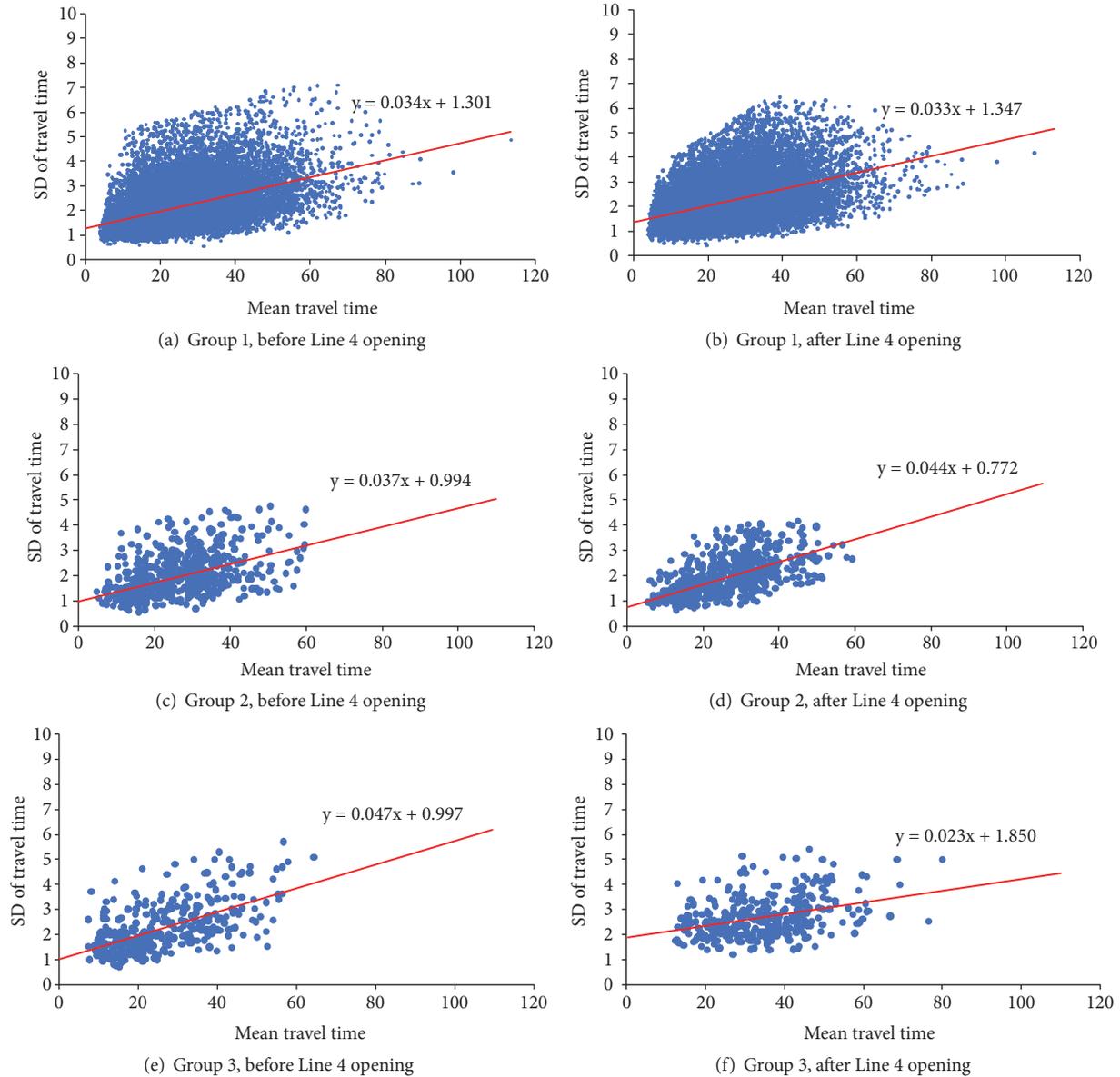


FIGURE 10: The mean and the SD of commuting travel time before and after Line 4 opening.

6. Conclusions

This paper presents a comprehensive analysis based on smart card data of metro system to demonstrate the effects of rail transit development in urban areas. The impacts of a new metro line (i.e., Line 4 of Nanjing metro) on both the whole metro network and different categories of commuters are investigated by analysing passenger flow distributions, travel times, and travel time reliability.

Analysis of passenger flow at all metro stations is conducted based on the smart card data before and after the opening of a new metro line. It is found that the passenger flow of the whole metro network increases after the new line opening, while the crowding at some metro stations reduced as the new line shares a large proportion of passengers.

To intensively study the impact of the new metro line on the network level, this study investigates the stations with significant change of passenger flow after the opening of the new metro line. Additionally, the travel time and the travel time reliability in the metro network are intensively studied. It is found that the opening of Line 4 leads to changes of travel time and travel time reliability for many stations.

To demonstrate the impact of the new metro line on different categories of commuters, metro commuters are identified and classified according to their OD choices. Flow distribution, travel time, and travel time reliability of different groups of commuters are analysed. The results indicate that the new metro line influences commuters with different OD choices to different degrees. The results in this study provide reference for guiding and evaluating rail transit network

planning in cities with large populations and have a potential to optimize and coordinate the operation between new transit lines and existing transit lines.

Several directions for further research are worth noting. First, although a big proportion of metro trips in Nanjing are made with smart card, it cannot be assumed that the flow pattern and travel times of passengers with one-way tickets are the same as smart card users. Due to the difficulty of collecting data of one-way tickets, the study presented in this paper can be extended by conducting activity and travel surveys. Second, in this study, smart card data from January to February 2017 are used to investigate the impact of the new metro line. In further research, more data after March 2017 are expected to be obtained to study passenger flow patterns and metro travel times to see how the new line contributes to the whole network after a long-term operation. Third, the analysis of travel time reliability presented in this paper can be extended to reliability-based accessibility by including detailed demographic information and land use data.

Data Availability

The data used to support the findings of this study can be found in the section of “Study Area and Data Collection”.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work described in this paper was jointly supported by the Natural Science Foundation of Jiangsu Province (BK20160676), the National Natural Science Foundations of China (71601045, 71501038, and 41571375), and the Social Science Foundation of Jiangsu Province (16GLC008).

References

- [1] M. Bagchi and P. R. White, “The potential of public transport smart card data,” *Transport Policy*, vol. 12, no. 5, pp. 464–474, 2005.
- [2] C. Chen, J. Ma, Y. Susilo, Y. Liu, and M. Wang, “The promises of big data and small data for travel behavior (aka human mobility) analysis,” *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 285–299, 2016.
- [3] C.-J. Cai, E.-J. Yao, S.-S. Liu, Y.-S. Zhang, and J. Liu, “Holiday destination choice behavior analysis based on AFC data of urban rail transit,” *Discrete Dynamics in Nature and Society*, vol. 6, pp. 1–7, 2015.
- [4] Y. Sun, J. Shi, and P. M. Schonfeld, “Identifying passenger flow characteristics and evaluating travel time reliability by visualizing AFC data: a case study of Shanghai Metro,” *Public Transport*, vol. 8, no. 3, pp. 341–363, 2016.
- [5] Y. Ma, W. Xu, X. Zhao, and Y. Li, “Modeling the hourly distribution of population at a high spatiotemporal resolution using subway smart card data: A case study in the central area of Beijing,” *ISPRS International Journal of Geo-Information*, vol. 6, no. 5, 2017b.
- [6] F. Zhou and R.-H. Xu, “Model of passenger flow assignment for Urban rail transit based on entry and exit time constraints,” *Transportation Research Record*, no. 2284, pp. 57–61, 2012.
- [7] J. Zhao, F. Zhang, L. Tu et al., “Estimation of Passenger Route Choice Pattern Using Smart Card Data for Complex Metro Systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 790–801, 2017.
- [8] S. Tao, J. Corcoran, F. Rowe, and M. Hickman, “To travel or not to travel: ‘Weather’ is the question. Modelling the effect of local weather conditions on bus ridership,” *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 147–167, 2018.
- [9] Y. Wang, G. H. D. A. Correia, E. D. Romph, and H. J. P. Timmermans, “Using metro smart card data to model location choice of after-work activities: An application to Shanghai,” *Journal of Transport Geography*, vol. 63, pp. 40–47, 2017.
- [10] W. Zhu, W. Wang, and Z. Huang, “Estimating train choices of rail transit passengers with real timetable and automatic fare collection data,” *Journal of Advanced Transportation*, vol. 2017, Article ID 5824051, 12 pages, 2017.
- [11] J. Kim, J. Corcoran, and M. Papamanolis, “Route choice stickiness of public transport passengers: Measuring habitual bus ridership behaviour using smart card data,” *Transportation Research Part C: Emerging Technologies*, vol. 83, pp. 146–164, 2017.
- [12] W. Jang, “Travel time and transfer analysis using transit smart card data,” *Transportation Research Record*, no. 2144, pp. 142–149, 2010.
- [13] X. Ma, Y. J. Wu, Y. Wang, F. Chen, and J. Liu, “Mining smart card data for transit riders travel patterns,” *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1–12, 2013.
- [14] H. Farooqi, M. Mesbah, and J. Kim, “Spatial-temporal similarity correlation between public transit passengers using smart card data,” *Journal of Advanced Transportation*, vol. 4, pp. 1–15, 2017.
- [15] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, “Spatio-Temporal Analysis of Passenger Travel Patterns in Massive Smart Card Data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–12, 2017a.
- [16] K. Kim, K. Oh, Y. K. Lee, S. Kim, and J.-Y. Jung, “An analysis on movement patterns between zones using smart card data in subway networks,” *International Journal of Geographical Information Science*, vol. 28, no. 9, pp. 1781–1801, 2014.
- [17] L. Sun and K. W. Axhausen, “Understanding urban mobility patterns with a probabilistic tensor factorization framework,” *Transportation Research Part B: Methodological*, vol. 91, pp. 511–524, 2016.
- [18] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, “Individual mobility prediction using transit smart card data,” *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 19–34, 2018.
- [19] C. Morency, M. Trépanier, and B. Agard, “Measuring transit use variability with smart-card data,” *Transport Policy*, vol. 14, no. 3, pp. 193–203, 2007.
- [20] S. Tao, D. Rohde, and J. Corcoran, “Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap,” *Journal of Transport Geography*, vol. 41, pp. 21–36, 2014.
- [21] L.-M. Kieu, A. Bhaskar, and E. Chung, “A modified density-based scanning algorithm with noise for spatial travel pattern analysis from smart card AFC data,” *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 193–207, 2015.
- [22] E. Manley, Z. Chen, and M. Batty, “Spatiotemporal variation in travel regularity through transit user profiling,” *Transportation*, vol. 45, no. 3, pp. 703–732, 2016.

- [23] L. M. Kieu, A. Bhaskar, and E. Chung, "Passenger segmentation using smart card data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1537–1548, 2015a.
- [24] X. Ma, C. Liu, H. Wen, Y. Wang, and Y. Wu, "Understanding commuting patterns using transit smart card data," *Journal of Transport Geography*, vol. 58, pp. 135–145, 2017.
- [25] J. Y. Park, D.-J. Kim, and Y. Lim, "Use of smart card data to define public transit use in Seoul, South Korea," *Transportation Research Record*, no. 2063, pp. 3–9, 2008.
- [26] Y.-S. Zhang and E.-J. Yao, "Splitting travel time based on AFC Data: Estimating walking, waiting, transfer, and in-vehicle travel times in metro system," *Discrete Dynamics in Nature and Society*, vol. 1, pp. 1–11, 2015.
- [27] M. Bernal, E. W. Welch, and P. S. Sriraj, "The effect of slow zones on ridership: An analysis of the Chicago Transit Authority "El" Blue Line," *Transportation Research Part A: Policy and Practice*, vol. 87, pp. 11–21, 2016.
- [28] J. Chan, *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data*. MS thesis, Massachusetts Institute of Technology, Cambridge, 2007.
- [29] D. Uniman, *Service reliability measurement framework using smart card data: application to the London underground*. MS thesis, Massachusetts Institute of Technology, Cambridge, 2009.
- [30] Z. Liu, Y. Yan, X. Qu, and Y. Zhang, "Bus stop-skipping scheme with random travel time," *Transportation Research Part C: Emerging Technologies*, vol. 35, pp. 46–56, 2013.
- [31] X. Fu, W. H. K. Lam, and B. Y. Chen, "A reliability-based traffic assignment model for multi-modal transport network under demand uncertainty," *Journal of Advanced Transportation*, vol. 48, no. 1, pp. 66–85, 2014.
- [32] Z.-L. Ma, L. Ferreira, M. Mesbah, and A. T. Hojati, "Modeling bus travel time reliability with supply and demand data from automatic vehicle location and smart card systems," *Transportation Research Record*, vol. 2533, pp. 17–27, 2015.
- [33] B. Y. Chen, H. Yuan, Q. Li et al., "Measuring place-based accessibility under travel time uncertainty," *International Journal of Geographical Information Science*, vol. 31, no. 4, pp. 783–804, 2017.
- [34] M. L. Senior, "Impacts on travel behaviour of Greater Manchester's light rail investment (Metrolink Phase 1): evidence from household surveys and Census data," *Journal of Transport Geography*, vol. 17, no. 3, pp. 187–197, 2009.
- [35] X. Cao and J. Schoner, "The influences of light rail transit on transit use: an exploration of station area residents along the Hiawatha Line in Minneapolis," in *Proceedings of the Transportation Research Board 92nd Annual Meeting*, 2013.
- [36] J. Harvey, N. Thorpe, M. Caygill, and A. Namdeo, "Public attitudes to and perceptions of high speed rail in the UK," *Transport Policy*, vol. 36, pp. 70–78, 2014.
- [37] J. Gadziński and A. Radzinski, "The first rapid tram line in Poland: How has it affected travel behaviours, housing choices and satisfaction, and apartment prices?" *Journal of Transport Geography*, vol. 54, pp. 451–463, 2016.
- [38] S. Saidi, S. C. Wirasinghe, and L. Kattan, "Long-term planning for ring-radial urban rail transit networks," *Transportation Research Part B: Methodological*, vol. 86, pp. 128–146, 2016.
- [39] A. Hong, M. G. Boarnet, and D. Houston, "New light rail transit and active travel: A longitudinal study," *Transportation Research Part A: Policy and Practice*, vol. 92, pp. 131–144, 2016.
- [40] S. Li, "Influence of a New Subway Line's Opening on Passenger Flow Characteristics of an Urban Rail Transit Network," in *Proceedings of the 15th COTA International Conference of Transportation Professionals: Efficient, Safe, and Green Multimodal Transportation, CICTP 2015*, pp. 1756–1769, China, July 2015.
- [41] S. Liu, E. Yao, and B. Li, "Exploring urban rail transit station-level ridership growth with network expansion," *Transportation Research Part D Transport & Environment*.
- [42] M.-P. Kwan, "Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility: Space-Time Integration in Geography and GIScience," *Annals of the Association of American Geographers*, vol. 103, no. 5, pp. 1078–1086, 2013.
- [43] B. Y. Chen, Y. Wang, D. Wang, Q. Li, W. H. K. Lam, and S. L. Shaw, "Understanding the impacts of human mobility on accessibility using massive mobile phone tracking data," *Annals of the American Association of Geographers*, vol. 108, pp. 1115–1133, 2018.
- [44] X. Fu and W. H. K. Lam, "A network equilibrium approach for modelling activity-travel pattern scheduling problems in multimodal transit networks with uncertainty," *Transportation*, vol. 41, no. 1, pp. 37–55, 2014.
- [45] K. M. N. Habib, L. Kattan, and T. Islam, "Model of personal attitudes towards transit service quality," *Journal of Advanced Transportation*, vol. 45, no. 4, pp. 271–285, 2011.
- [46] H. Wakabayashi and Y. Matsumoto, "Comparative study on travel time reliability indexes for highway users and operators," *Journal of Advanced Transportation*, vol. 46, no. 4, pp. 318–339, 2012.

Research Article

Improving Traffic State Prediction Model for Variable Speed Limit Control by Introducing Stochastic Supply and Demand

Yuwei Bie ¹, Mudasser Seraj,¹ Can Zhang,¹ and Tony Z. Qiu ^{1,2}

¹Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada

²Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan, China

Correspondence should be addressed to Tony Z. Qiu; zhijunqiu@ualberta.ca

Received 19 March 2018; Revised 6 July 2018; Accepted 24 July 2018; Published 5 August 2018

Academic Editor: David Z. W. Wang

Copyright © 2018 Yuwei Bie 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.

Variable speed limit (VSL) is becoming recognized as an effective way to improve traffic throughput and road safety. In particular, methods based on traffic state prediction exhibit promising potential to prevent future traffic congestion and collisions. However, field observations indicate that the traffic state prediction model results in nonnegligible error that impacts the next step decision making of VSL. Thus, this paper investigates how to eliminate this prediction error within a VSL environment. In this study, the traffic state prediction model is a second-order traffic flow model named METANET, while the VSL control is model predictive control (MPC) based, and the VSL decision is discrete optimized choice. A simplified version of the switching mode stochastic cell transmission model (SCTM) is integrated with the METANET model to eliminate the prediction error. The performance of the proposed method is assessed using field data from a VSL pilot test in Edmonton, Canada, and is compared with the prediction results of the baseline METANET model during the road test. The results show that during the most congested period the proposed SCTM-METANET model significantly improves the prediction accuracy of regular METANET model.

1. Introduction

The VSL control method was designed to keep the credibility of speed limits even under adverse conditions such as congestion and bottleneck road segments, so that the speed limits still work to maintain traffic safety and the highest possible traffic throughput. This method adjusts the posted speed limit based on real-time road, traffic, and weather conditions [1]. The VSL control method performs well in terms of control scale, control method flexibility, and feasibility. The most commonly used VSL strategies take a reactive approach with simple policies: using this strategy, the VSL decision is triggered by abnormal traffic states detected in real time, such as high traffic flow or low average speed. However, to achieve improved traffic throughput, a more complex control algorithm that includes optimized VSL decision-making is needed. Proactive VSL control goes one step further than reactive VSL control since it is optimized and calculated based on the predicted traffic state, so as to prevent a severe traffic breakdown from happening. Between two general patterns

of VSL control, stimulation-reactive pattern and proactive pattern, this paper considers the latter pattern to be more challenging but promising.

In proactive VSL control, an accurate macroscopic level traffic prediction model is the core part. Over the last decades, various traffic prediction models were invented: the author classifies them to statistical and nonstatistical methods [2]. The statistical methods can predict the traffic flow with high accuracy under stable traffic conditions, commonly used methods including the class of Time Series models, such as seasonal Autoregressive Integrated Moving Average model [3] and Kalman Filter State-Space model [4], Neural Network [5], and Nonparametric Regression [6]. However, in the case of VSL control, traffic flow prediction is required under congestion condition at bottleneck road sections where traffic conditions can change abruptly, so that the author chose non-statistical methods, i.e., the traffic prediction model derived from traffic flow models. Those models have potential to reflect the nature of traffic flow and capture the fluctuation of traffic flow: the traffic flow is treated as fluid that is aggregated

by individual vehicles, and three variables are capable of describing the characteristics of traffic stream; they are flow, density, and mean speed. The traffic flow models are classified as first-, second-, or higher-order models, depending on the number of the included differential equations [7]. First-order model is able to describe traffic flow, but not traffic density or speed. Among all first-order models, the most used is Lighthill–Whitham–Richards model [8, 9] using one partial differential equation to describe the vehicle flow conservation law. This was also the first time that traffic flow model is combined with static fundamental diagram. Another representative first-order model is the Cell Transmission Model [10–12], which is a discretized and simplified version of the Lighthill–Whitham–Richards model. The Payne model [13] is the oldest second-order traffic flow model. Besides the flow conservation law equation, the Payne model also includes one partial differential equation that describes mean speed dynamics. This model can replicate traffic phenomenon with higher accuracy. There are also other types of second-order traffic prediction models and combinations, such as variation kinematic waves [14], second-order traffic flow model with Kalman filter [15], and CTM-based second-order traffic flow model with particle filtering [16].

To achieve the highest traffic prediction accuracy, in this paper, a second-order traffic flow model named METANET [17, 18] is chosen. It is a discretized and enhanced version of the Lighthill–Whitham–Richards model combined with the Payne model. As it is a second-order traffic flow model, it is able to predict traffic density, average vehicle speed, and traffic flow by three dynamic functions in the model. METANET model fulfills the simplicity and convenience requirement since it has a space-time discrete, explicit analytical state-space form and allow for convenient discretization intervals [7]. What is more, METANET model is proved by previous studies that it is of high accuracy and relatively easy to calibrate, which makes it one of the most frequently utilized macroscopic traffic flow models in variety of traffic engineering tasks and researchers, and this model is capable of being used for optimal real-time traffic control of freeway traffic.

However, due to the stochastic nature of traffic flow, accurately predicting the traffic state of bottlenecks for VSL control is a complex task, especially during peak hours and for congested road segments. The traffic flow, speed, and density are by nature stochastic and variable, when traffic flow models are used for predicting traffic condition for making VSL control decisions, those stochasticity and variability need to be described and modeled and be added to the original static model to meet the specific high prediction requirement of VSL control. VSL field experiments in this paper have demonstrated that the traffic state prediction model exhibits larger error during the afternoon peak period. The stochastic nature of traffic can be modeled by adding an exogenous disturbance term to each round of calculation. This paper is inspired by the stochastic elements in the stochastic cell transmission model (SCTM) [19] and incorporates them in this study. The most straightforward method to model the stochasticity is to apply a sequential Monte Carlo simulation

method to the CTM to mimic the change from free flow traffic conditions to congested conditions [20], but this method has a high computation cost. Another way to incorporate the stochastic element is to set the sending and receiving flow as random variables [21]. However, this method does not function well when traffic is light or congested, for the random terms cannot represent transportation features. In 2011, Sumalee et al. [19] proposed a stochastic CTM (SCTM) to describe the macroscopic dynamics of traffic flow under demand and supply uncertainties. Its switching mode design simplifies the sequential Monte Carlo method, and its stochastic element takes into account the fundamental diagram parameters instead of purely random generated. The SCTM has been implemented in practical research projects, such as dynamic journey time estimation [22] and road network traffic assignment under adverse weather [23]. The cell transmission model (CTM) is a first-order traffic flow model with no exogenous input, and it serves as the density dynamic function in the METANET model. In this paper, the regular density dynamic function in METANET model, also called CTM is replaced with the SCTM. Stochastic elements are added in anticipation of better prediction performance.

Besides traffic prediction model, the VSL control structure also has various choices [24] such as case-based reasoning, knowledge-based systems (in which the policy is based on previous practices), and model-based prediction. To find the optimal control structure that utilizes the traffic state prediction model, a MPC framework [25] is applied in this paper, forming a closed loop. The future traffic state is predicted, and then the optimal VSL decision is made based on prediction results. MPC is an optimal control scheme applied within a rolling horizon framework. Optimal control was successfully applied by Kotsialos et al. [26] to coordinate or integrate traffic control measures. Both optimal control and MPC have the advantage of a controller that generates an optimal control signal or decision according to a user-defined objective function. The objective function in this paper is a weighted summation of total travel time (TTT) and total travel distance (TTD), where the total travel time is minimized and total travel distance is maximized simultaneously.

VSL control has been successfully implemented in European countries and the United States [27]. In Edmonton, Canada, a four-week VSL field test was implemented on a 10-kilometer stretch of freeway, and this field test serves as a case study in this paper. The field data truly reflect the driver behavior and compliance of people in a major western Canadian city. The actual field data and METANET-predicted traffic data are extracted for comparison with the simulated data generated by the proposed model. This paper first integrates the SCTM to the METANET model for VSL control purposes and then tests its feasibility using both field experiment data and computer simulation data. The upgrading of METANET model, the field test result analysis, and the simulation provide references for future local researches.

The following sections of this paper include a literature review of VSL and the original SCTM model; an explanation

of the model predictive control- (MPC-) based VSL algorithm; a description of the modified SCTM that is used in this study; presentation of the case study; and, finally, a discussion of the results and conclusion.

2. A Model Predictive Control Based Variable Speed Limit Algorithm

The model description is separated into four parts: the description of traffic state prediction model and the description of fundamental diagram used in this study. Section 2.1 will be notations and assumptions. Section 2.2 will describe in detail the baseline VSL control logic and algorithm. Section 2.3 will present model calibration skills including the calibration of fundamental diagram and the macroscopic traffic prediction model involved in VSL control.

2.1. Notations and Assumptions. For consistency, the freeway is divided into N sections with lengths Δ_i , $i = 1, \dots, N$, each having at most one on-ramp and off-ramp. All the variables used throughout this paper are defined as follows:

T : data collection interval=20s

λ_i : number of lanes at section i

Δ_i : length of section i

$q_i(k)$: number of vehicles in the freeway section i at time $k \cdot T$ divided by the length Δ_i

$\rho_i(k)$: traffic density at time step k , section i .

$v_i(k)$: average vehicle travel speed at time step k , section i .

$u_i(k)$: the optimized speed limit posted at time step k , section i .

$r_i(k)$: on-ramp flow at time step k , section i

$s_i(k)$: off-ramp flow at time step k , section i

$V(\rho_i(k))$: desire speed in speed dynamics in METANET model

Θ : set of unknown parameters of METANET model

$\rho_{cr,i}$: critical density of FD at section i

$\rho_{jam,i}$: jam density of FD at section i

$v_{f,i}$: free flow speed of FD at section i

w_i : shockwave speed of FD at section i

θ : capacity drop fraction of FD at section i

TTT: total travel time

TTD: total travel distance

α_{TTT} : weight of TTT

α_{TTD} : weight of TTD.

2.2. The MPC-Based VSL Algorithm Using METANET Prediction Model. The MPC-based VSL algorithm is composed of five parts: an original traffic prediction model, control case traffic prediction model, optimizer, objective function, and constraints. As shown in Figure 1, for VSL control road

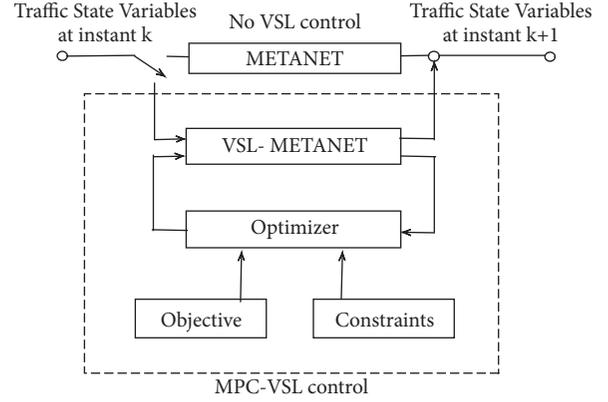


FIGURE 1: The block diagram of model predictive VSL control.

segments, the system uses a modified METANET model for traffic state prediction, and the discrete choice optimizer makes a choice based on predicted traffic states, given the objective function and constraints. After optimization, the optimal VSL choice is shown to operators, and, simultaneously, the VSL value is inputted into the VSL-METANET model again for the final traffic state prediction results to be recorded. Figure 1 shows a block diagram of the VSL control system.

The METANET model as a second-order traffic flow model introduces speed dynamic functions instead of including only density dynamics to describe flow conservation law. This feature of the METANET model enables it to predict density, speed, and flow variables, respectively, accurately, as well as with a small time interval. The discretized model formulation of the original METANET is written as follows.

Density dynamics (also referred to the regular CTM) are as follows:

$$\begin{aligned} \rho_i(k+1) &= \rho_i(k) \\ &+ \frac{T}{\Delta_i \lambda_i} [\lambda_{i-1} q_{i-1}(k) - \lambda_i q_i(k) + r_i(k) - s_i(k)] \end{aligned} \quad (1)$$

Speed dynamics are as follows:

$$\begin{aligned} v_i(k+1) &= v_i(k) + \frac{T}{\tau} [V(\rho_i(k)) - v_i(k)] \\ &+ \frac{T}{\Delta_i} v_i(k) [v_{i-1}(k) - v_i(k)] \\ &- \frac{\eta \cdot T [\rho_{i-1}(k) - \rho_i(k)]}{\tau \cdot \Delta_i \cdot [\rho_i(k) + \kappa]} \end{aligned} \quad (2)$$

Flow dynamics are as follows:

$$q_i(k+1) = \rho_i(k+1) \cdot v_i(k+1) \quad (3)$$

where i is the index of links, and $i = 1, 2, \dots, M$, with M representing the number of testbed sections. K is the index of time instants, and T is the calculating time interval, where

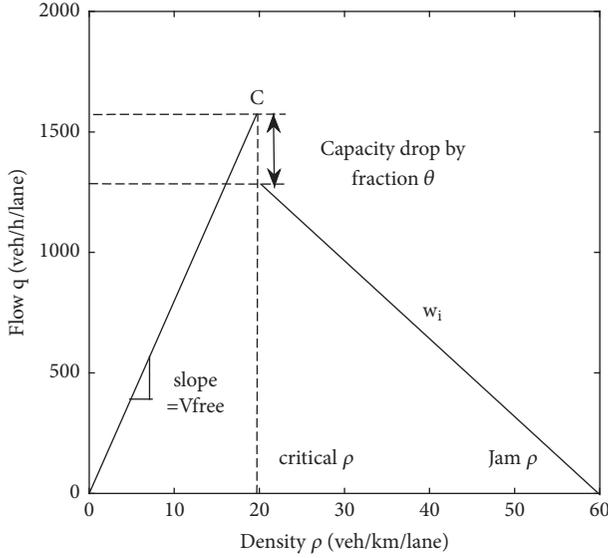


FIGURE 2: The triangular FD assumption.

$T=20s$. Δ_i represents the segment length of link i , and λ is the lane number of link i . In the model, τ , κ , α , and η (km^2/h) are global model parameters calibrated using the historical data. The desired speed $V(\rho_i(k))$ (km/h) in the speed dynamics is represented by the following expression:

$$V(\rho_i(k)) = v_{f,i} \exp \left[-\frac{1}{\alpha} \left(\frac{\rho_i(k)}{\rho_{c,i}} \right)^\alpha \right] \quad (4)$$

The triangular fundamental diagram (FD) is assumed in the control algorithm. Figure 2 shows the triangular FD assumption used in this paper. In (4), $v_{f,i}$ is free flow speed and $\rho_{c,i}$ represents critical density, which is the density associated with capacity. In this model, the density dynamics do not involve parameters, and other parameters in the speed dynamics are calibrated globally for goodness of fit. Of the four terms that make up the speed dynamics equation, each term has a physical meaning. In (2), the second term is referred to as the relaxation term, indicating that, with a lag time item τ , the mean speed v of the link “relaxes” to the desired speed, which largely depends on the parameters of the FD. The selection of the desired speed is critical in reflecting driver behavior, and, based on previous practice [28], the format of (4) was chosen. The third term is the convection term used to indicate that vehicles traveling from the upstream link $i-1$ to the current link i adapt their speed gradually rather than instantaneously. The fourth term is the anticipation term, indicating that drivers are always watching the road ahead for changes in traffic. If a driver observes high traffic density in the downstream link $i+1$, the driver then reacts by slowing down. The constant $\kappa > 0$ is added to keep the anticipation term limited when density is low.

The METANET algorithm for the VSL control environment is modified by replacing the desired speed with a modified one. In the control segments, the desired speed is assumed to be the posted advisory speed, rather than the

speed determined by the fundamental diagram as in the non-control case. In the VSL-METANET model, the desired speed term then becomes the optimized VSL value, as decided by the optimizer in the MPC system.

Desired speed term when VSL is triggered is as follows:

$$V(\rho_i(k)) = u_i(k) \quad (5)$$

The objective function minimizes a weighted sum of TTT and TTD in a discrete choice fashion. In terms of definition, TTT indicates a weighted sum of density of all time steps and links. Only minimizing TTT produces system bias, wherein the optimizer tends to choose a lower speed limit to decrease the density, so that TTD is maximized simultaneously. In the objective function, weights α_{TTT} and α_{TTD} are applied to TTT and TTD.

$$TTT = T \sum_{j=1}^{pr-1} \sum_{i=1}^M \lambda_i L_i \rho_i(k+j) \quad (6)$$

$$TTD = T \sum_{j=1}^{pr-1} \sum_{i=1}^M \lambda_i L_i \rho_i(k+j) v_i(k+j) \quad (7)$$

$$J = T \sum_{j=1}^{pr-1} \sum_{i=1}^M \lambda_i L_i [\alpha_{TTT} \rho_i(k+j) - \alpha_{TTD} \rho_i(k+j) v_i(k+j)] \quad (8)$$

The desired speed under the VSL environment is decided by the optimizer. The optimization is subjected to several constraints of traffic safety, driver acceptance, and traffic flow characteristics [1]. The first constraint is the upper limit of $u_i(k)$, since the regular speed limit is 80 km/h , under congested conditions $u_i(k) \leq 80$. The second constraint is to maintain continuous traffic flow even under congestion. From local regulations, the lower bound of the speed limit is 30 km/h , which means $u_i(k) \geq 30$. In Canada, the speed limit is in multiples of 10 km/h ; therefore, to ensure safety, the change in speed limit between two time instants is 0 or 10 km/h , represented as $u_i(k) - u_i(k+1) \in \{-10, 0, 10\}$. The optimizer then becomes a discrete choice model.

2.3. Calibrating the FD and METANET Model. The parameter estimation of the FD is based on data collected by conventional loop detectors. For the parameter calibration of the FD, free flow speed $v_{f,i}$, critical density $\rho_{cr,i}$, capacity C_i , jam density $\rho_{jam,i}$, and capacity drop fraction θ must be estimated. For calibration, the data format used is (ρ_i, q_i) data points, where the horizontal coordinate is density and the vertical coordinate is flow, as Figure 2 shows. The procedure for obtaining the necessary parameters from a triangular FD is as follows.

(1) The identification of capacity C_i : in defining the triangular FD, the summit point of the triangle roughly indicates critical density and capacity. All (ρ_i, q_i) points are plotted to find the third largest q_i and then take it as capacity C_i . The largest flow is not chosen because it is normally an extreme outlier; this could be due to detection error so that the third largest flow is empirically chosen as capacity.

(2) The identification of critical density $\rho_{cr,i}$: after determining C_i , the corresponding traffic density ρ_i is then identified as critical density $\rho_{cr,i}$. This identification is based on the assumption of triangular shape FD: the vertical line across C_i intersects the horizontal density axis at critical density $\rho_{cr,i}$. Point $(\rho_{cr,i}, C_i)$ is then the vertex of the triangular FD.

(3) The identification of free flow speed $v_{f,i}$: after defining $\rho_{cr,i}$, the whole dataset can be divided into two parts, the left and right side of the triangle. The left side of the triangle represents uncongested traffic conditions and the right side of the triangle represents congested traffic conditions. Then the slope of each data point distributed on the left side is calculated, and the average is taken as $v_{f,i}$, as shown in the following equation [2]:

$$v_{f,i} = \frac{1}{n} \left(\sum_{ii=1}^n \frac{q_{ii}}{\rho_{ii}} \right), \quad \rho_{ii} \in (0, \rho_{cr,i}) \quad (9)$$

(4) The identification of jam density $\rho_{jam,i}$, the capacity drop fraction θ , and shockwave speed w_i : the jam density $\rho_{jam,i}$ represents a theoretical value that indicates when the road section is totally congested and all vehicles have stopped moving. From the dataset, the data point with the largest density value and zero speed is selected as $\rho_{jam,i}$. The slope of the right side, i.e., shockwave speed w_i is determined by least square fit fixing one foot $(\rho_{jam,i}, 0)$. After determining one foot and the slope of the right side, the intersecting point of the regression line and the vertical line across C_i is the congested capacity after a capacity drop fraction of θ .

For the parameter estimation of METANET model, τ , κ , α , and η (km²/h) are global model parameters to be calibrated. The calibration of parameters is not done in real time, according to the logic that both parameters from FD and METANET model represent aggregated traffic behavior of a period of time, the calibration happens right before the field test and uses traffic data collected in the last 7 days. In doing SCTM-METANET simulation, the same parameter set as in the field test is used. The parameters of METANET model are estimated via the following expression:

$$\Theta^* = \arg \min_{\Theta} \left[\sum_{ii}^N \left\{ v_{obs}^{ii} - f(v_{obs}^{ii} | \Theta) \right\}^2 + \sum_{ii}^N \left\{ \rho_{obs}^{ii} - f(\rho_{obs}^{ii} | \Theta) \right\}^2 \right] \quad (10)$$

3. Integrating Switching-Mode SCTM with METANET Model

The switching-mode SCTM is employed to replace the regular CTM in the METANET model. The traditional CTM, derived from the conservation law of traffic flow, is a single linear system without exogenous input or tuning parameters. When CTM is used for traffic state prediction, after rounds of iteration, the model is farther and farther from reality since during the calculation time the outside traffic condition changes quickly. In this paper, the traffic prediction length

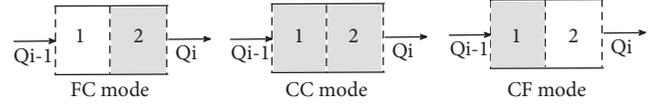


FIGURE 3: Traffic operation modes.

in the VSL field test is 5 minutes, containing 15 iterations of calculation, but congestion can form within 3 minutes. To solve this problem, the SCTM introduces exogenous stochastic disturbance terms to each iteration of computing to counteract the prediction error.

Incorporating stochastic supply and demand into a traffic flow model can be different ways, the FD stochasticity, the boundary condition stochasticity, and stochastic adjusting parameter if necessary. In the case of the SCTM in this paper, the stochastic elements are added in two ways: (1) demand stochasticity, the inflow disturbance term added to the CTM; (2) supply stochasticity also FD stochasticity, each of the disturbance terms containing stochastic FD parameters extracted by statistical technics.

Another feature of the SCTM is that, under different traffic condition modes, the disturbance elements change slightly. The original SCTM has five modes (refer to [19]), and in this paper the number of modes is simplified to three because the CTM is still used for free flow conditions and the wave front of congestion is only from downstream to upstream. The three modes are free flow to congestion (FC) mode, congestion to congestion (CC) mode, and congestion to free flow (CF) mode. Different from the CTM, the SCTM considers the adjacent two cells as a whole and uses vector $P(k) = [\rho_1(k), \rho_2(k)]^T$ in the density dynamics calculation. Please see Figure 3 for the illustration of traffic operation modes where white segment represents for low traffic density free flow condition and grey segment represents for high traffic density congested condition.

When defining the traffic operational modes, a typical SCTM develops complicated probabilistic occurrence models as mode switches [19], and it serves generalized traffic situations. However, when SCTM is implemented in a specific VSL control road segment whose traffic condition is highly repeatable and well understood, the probability expression then decays to a fixed version. The fixed mode switch time points are empirically set based on traffic data, which can be observed in the case study of this paper. Additionally, the switch time considers the suggestion of local authorities. Free flow mode, which is the original density dynamic function in METANET, is in effect 4:00pm to 4:59pm. The FC mode takes effect during 5:00pm to 5:09pm and serves as a transitional stage between free flow mode and congested mode. The CC mode takes place between 5:10pm and 5:49pm. Then the CF mode takes effect between 5:50pm and 5:59pm. From 6:00pm on, the CTM takes over again. In the FC mode, $u(k) = (q_{i-1}(k), r_i(k), s_i(k), q_i(k))^T$, $\omega_2(k) = w_{c,2}(k)$, and $\lambda(k) = (0, \rho_{jam,2}(k))^T$.

Density dynamics (SCTM) under FC mode are

$$P(k+1) = (I + A_2 \omega_2(k)) P(k) + B_2 \omega_2(k) \lambda(k) + Bu(k) \quad (11)$$

where

$$\begin{aligned} A_2 &= \begin{bmatrix} 0 & \frac{T_s}{l_1} \\ 0 & -\frac{T_s}{l_2} \end{bmatrix}, \\ B_2 &= \begin{bmatrix} 0 & -\frac{T_s}{l_1} \\ 0 & \frac{T_s}{l_2} \end{bmatrix}, \\ B &= \begin{bmatrix} \frac{T_s}{l_1} & -\frac{T_s}{l_1} & 0 & 0 \\ 0 & 0 & -\frac{T_s}{l_2} & \frac{T_s}{l_2} \end{bmatrix}. \end{aligned} \quad (12)$$

After the traffic condition transitions to congestion mode, the CC model applies. In the CC model, $\omega_i(k) = w_i(k)$, and $\lambda(k) = (\rho_{jam,1}(k), \rho_{jam,2}(k))^T$.

Density dynamics (SCTM) under CC mode are

$$\begin{aligned} P(k+1) &= \left(I + \sum_{i=1}^2 A_i \omega_i(k) \right) P(k) \\ &\quad + \sum_{i=1}^2 B_i \omega_i(k) \lambda(k) + Bu(k) \end{aligned} \quad (13)$$

where

$$\begin{aligned} A_1 &= \begin{bmatrix} -\frac{T_s}{l_1} & 0 \\ 0 & 0 \end{bmatrix}, \\ A_2 &= \begin{bmatrix} 0 & \frac{T_s}{l_1} \\ 0 & -\frac{T_s}{l_2} \end{bmatrix}, \\ B &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{T_s}{l_2} & -\frac{T_s}{l_2} \end{bmatrix}, \\ B_i &= -A_i, \quad i = 1, 2. \end{aligned} \quad (14)$$

After the traffic condition transitions out of congestion mode, the CF model applies. In the CF model, $\omega_1(k) = w_1(k)$, $\omega_2(k) = v_{f,2}(k)$, and $\lambda(k) = (\rho_{jam,1}(k), C_1(k))^T$.

Density dynamics (SCTM) under CF mode are

$$\begin{aligned} P(k+1) &= \left(I + \sum_{i=1}^2 A_i \omega_i(k) \right) P(k) \\ &\quad + \left(B_0 + \sum_{i=1}^2 B_i \omega_i(k) \right) \lambda(k) + Bu(k) \end{aligned} \quad (15)$$

where

$$\begin{aligned} A_1 &= \begin{bmatrix} -\frac{T_s}{l_1} & 0 \\ 0 & 0 \end{bmatrix}, \\ A_2 &= \begin{bmatrix} 0 & 0 \\ 0 & -\frac{T_s}{l_2} \end{bmatrix}, \\ B_0 &= \begin{bmatrix} 0 & -\frac{T_s}{l_1} \\ 0 & \frac{T_s}{l_2} \end{bmatrix}, \\ B_1 &= -A_1, \quad B = B_2 = 0. \end{aligned} \quad (16)$$

4. Case Study

The vehicle detection stations (VDS) on a 10-kilometer corridor of Whitemud Drive in Edmonton, Canada, collect, and store traffic data from dual loop detectors. This section of road plays an important role in people and freight transportation in the city. The VDS system currently has 28 VDS in total, and each station has three or four dual loops. The data recording frequency is 20 seconds. Each dual loop reports the volume q —the number of vehicles crossing the loop detector during a 20-second time interval—and mean speed measurement v , and the occupancy measurement, which cannot be used directly in traffic control and cannot be transformed accurately into density. The accurate density measurement ρ is calculated by $\rho = q/v$. Figure 4 shows the testbed geometry, VDS locations and the location of dynamic message signs (DMS) that will be used to display variable speed limits.

The VSL control decisions are made and sent out by software with the METANET model embedded. When the detected average speed is below the posted speed of 80 km/h, the VSL-METANET prediction is triggered on the corresponding road segment. The prediction step length is 20 seconds and 15 iterations, which means the prediction horizon is 5 minutes. The prediction is calculated every minute. So that all variables for comparison are 1 min aggregated, and when simulating SCTM-METANET the same aggregation level is used. Theoretically speaking, the METANET model can be made to predict farther future such as 10 min and 15 min; however, in real life situation in Edmonton, Canada especially during peak hours, traffic condition changes drastically even within 5 minutes (please see Figures 6(b), 6(c), and 6(d) and Figures 7(b), 7(c), and 7(d)). So that when doing municipal level experiment on freeways, prediction horizon longer than 5 minutes was rarely used. If under non-peak hours when traffic flow is relatively steady, it is possible to recalibrate METANET model and achieve high accuracy prediction for longer horizons.

During real-time VSL control, data from loop detector stations are sent to computers in the traffic management center in the City of Edmonton. The software installed in the computers is responsible for providing speed limit

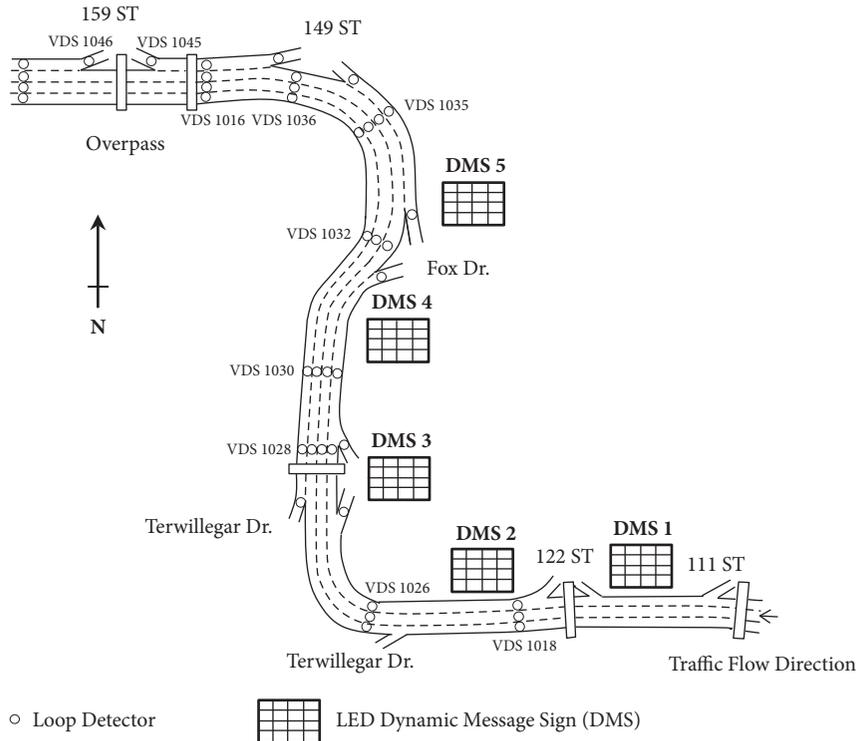


FIGURE 4: Vehicle detection stations and dynamic message signs on Whitemud Drive, Edmonton, Canada.

suggestions. The decisions are then shown on screen with beeping notifications for human operators to check and send to DMS on the road. From August 13 to September 4 of 2015 on weekdays, the City of Edmonton conducted a four-week VSL pilot test on westbound Whitemud Drive. Five DMSs in total were located upstream of predefined congestion-prone points. From the testing period, DMS 1 on August 17 was chosen as a case for this study because a period of normal peak hour congestion occurred and the VSL functioned well. Due to legal reasons, the VSLs were not mandatory to drivers, and they were labeled as “Advisory Driving Speed” on the DMSs.

The performance measurements of VSL in this study are TTT and TTD, which can be taken together as the measure of effectiveness (MOE), as shown in (6) and (7). At each time step, the TTT and TTD are calculated from a weighted sum of the density and flow of all links, in which the weight is the segment length of each link. Figure 5 shows the speed profile of the test day. VSL is observed to help stabilize the speed and the predefined fixed time FC-CC-CF mode switch makes sense in this case.

In the field test, the global parameters of METANET model were calibrated based on historical traffic data of August 2014, and, during field test, those parameters were fixed. The reason for fixing global parameter is to reduce the complexity of online traffic prediction. Table 1 shows the precalibrated global parameters of METANET model.

4.1. The Measurement of Prediction Error and the Performances of SCTM-METANET. The discrepancies between the actual

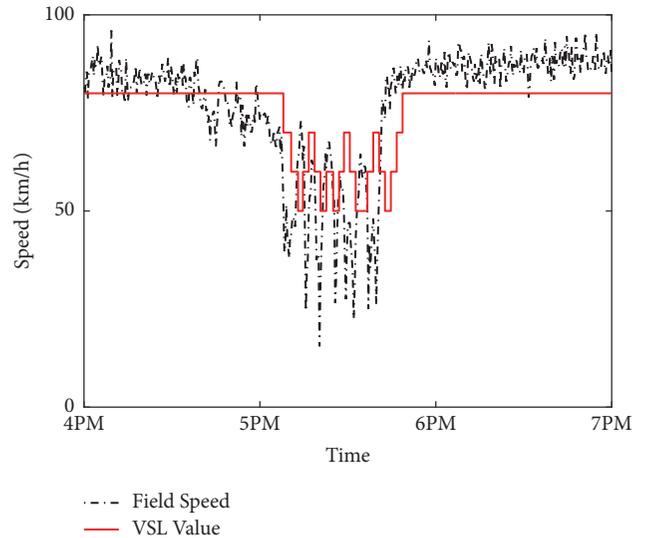


FIGURE 5: Speed profile of case study day Aug.17, 2015.

MOE and those predicted can be categorized as structured error and unstructured error. Structured error occurs when fixed parameters, such as segment lengths, are incorrectly measured. Unstructured error is the true error or the error that is found in the prediction model, which usually happens after eliminating structured error.

The structured error (SE) of the predicted MOE is assumed to be proportional to the original predicted MOE, after denying other assumptions such as proportion to

TABLE 1: Parameters used in the field test.

METANET Parameters	η	κ	α	τ
	37.98	10.00	2.29	120.00
FD Parameters	w_1, w_2	$v_{f,1}, v_{f,2}$	$\rho_{jam,1}, \rho_{jam,2}$	
	12	82	118	

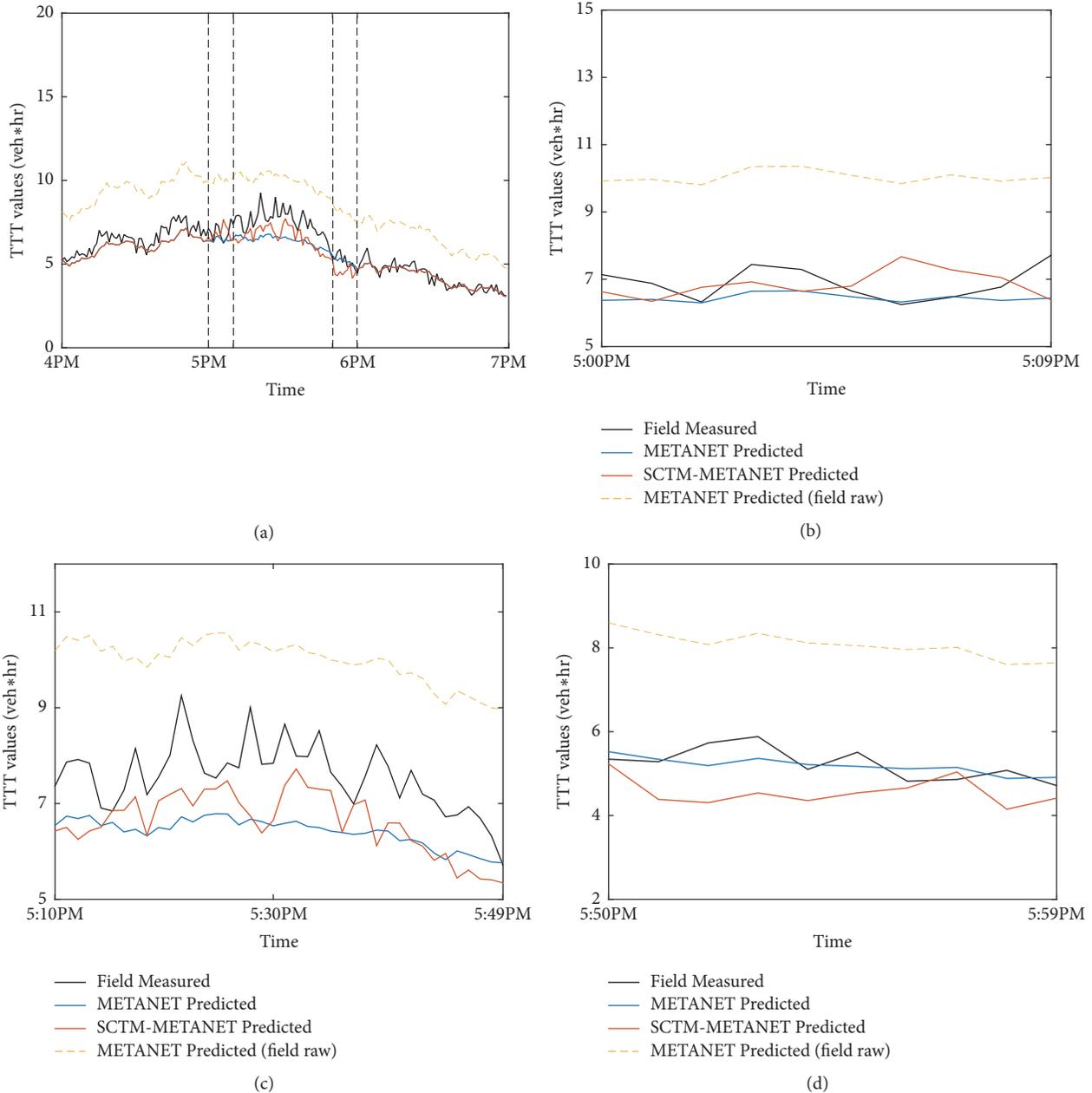


FIGURE 6: (a) Comparing field-measured TTT, original METANET, and SCTM-METANET predicted TTT; (b), (c), (d) detailed view of FC, CC, and CF modes.

measured MOE, fixed value of error, and normal distribution. So that the adjusted predicted MOE is calculated by scaling down original predicted MOE. The reference point is the beginning of the study period 4:00:00PM, meaning the suggestion that at the start point of this time period

no structured prediction uncertainty existed. Based on the above understanding, the overall structured uncertainty can be calculated. In Figures 6 and 7, SE refers to the distance between the curve METANET-predicted (field raw) and the curve METANET-predicted.

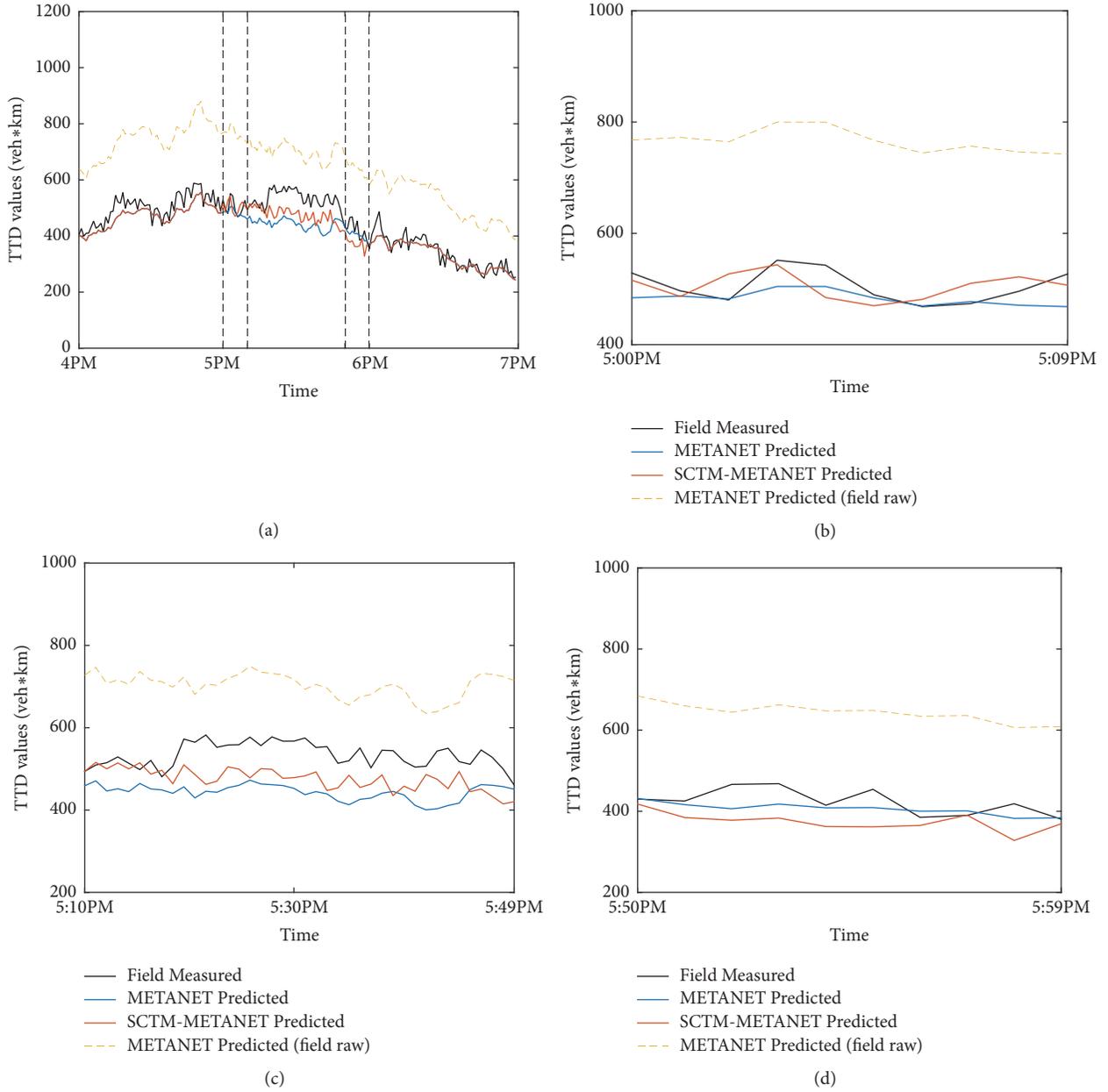


FIGURE 7: (a): Comparing field-measured TTD, original METANET, and SCTM-METANET predicted TTD. (b), (c), (d) detailed view of FC, CC, and CF modes.

$$SE(t) = \left(1 - \frac{Moe_m(t')}{Moe_p(t')} \Big|_{t'=4:00:00PM} \right) \cdot Moe_p(t) \quad (17)$$

where $Moe_m(t)$ and $Moe_p(t)$ represent field-measured and METANET-predicted TTT and TTD, respectively, at time step t . After eliminating structured error by scaling down the original predicted MOE, unstructured error (UE) remains, which represents the true performance of the prediction model. In Figures 6 and 7, unstructured error (UE) refers to the difference between the curve METANET-predicted and field-measured. In the following section, the

root mean square error (RMSE) of both the original and new models is calculated from the UE.

$$UE(t) = Moe_m(t) - Moe_p(t)$$

$$\frac{Moe_m(t')}{Moe_p(t')} \Big|_{t'=4:00:00PM} \quad (18)$$

Figures 6 and 7 and Tables 2 and 3 demonstrate the performance of the proposed SCTM-METANET versus the regular METANET. The raw field data predicted by METANET are collected from the field control software, and, after eliminating SE, the data become METANET-predicted data,

TABLE 2: Quantitative comparison of METANET and SCTM-METANET models: TTT.

RMSE of TTT prediction	Free Flow (4:00-4:59pm)	FC Mode (5:00-5:09pm)	CC Mode (5:10-5:49pm)	CF Mode (5:50-5:59pm)	Free Flow (6:00-6:59pm)
Conventional METANET	0.40	0.61	1.27	0.31	0.19
SCTM-METANET	— —	0.77	1.09	0.84	— —
Improvement (Percentage)	0	-26%	14%	-171%	0
No. of Data Points	60	10	40	10	60

TABLE 3: Quantitative comparison of METANET and SCTM-METANET Models: TTD.

RMSE of TTD prediction	Free Flow (4:00-4:59pm)	FC Period (5:00-5:09pm)	CC Period (5:10-5:49pm)	CF Period (5:50-5:59pm)	Free Flow (6:00-6:59pm)
Conventional METANET	20.0	31.42	95.97	31.53	25.5
SCTM-METANET	— —	29.76	67.96	60.69	— —
Improvement (Percentage)	0	5%	29%	-92%	0
No. of Data Points	60	10	40	10	60

which are ready to be compared with the computer-simulated SCTM-METANET.

Figure 6 shows that during free flow before 5PM and after 6PM the regular METANET model satisfactorily captures the trend of traffic flow. During 5:10PM and 5:49PM, the most congested period, the original METANET largely underestimates the TTT as well as its fluctuation. After introducing the SCTM into METANET, the predicted TTT not only is closer to reality, but also better matches the oscillating trend of the field-detected TTT. Figures 6(a) and 6(c) show steady improvement of the proposed SCTM-METANET model. Under transitional mode FC and CF, the SCTM-METANET does not show visible improvement of prediction accuracy, which may be due to too few data points. The FC and CF mode last only for 10 minutes, which means each condition has only 30 data points for computation. In summary, the SCTM-METANET outperforms the regular METANET model with noticeable improvements in stability, especially during the most congested period. Those improvements can be observed not only in the figure, but also quantitatively in Table 2.

When focusing on the prediction error during the CC period, the proposed SCTM-METANET model shows a significant 14% improvement calculated from 40 data points. Quantitative comparison also shows that, during the FC mode and CF mode, the proposed model shows unstable performance with negative improvements. The free flow period before and after SCTM-METANET shows stable and best prediction accuracy with the lowest error.

In the prediction of TTD, the SCTM-METANET performs better than for TTT. Figure 7 shows that, during free flow hours before 5PM and after 6PM, the regular METANET model satisfactorily captured the traffic flow trend. Also, during 5:10PM and 5:49PM, the original METANET largely underestimated the TTD and its fluctuation. When using the SCTM-METANET, the predicted TTD was not only closer to reality, but also better matched the oscillating trend of the field-detected TTD. Figures 7(a) and 7(c)

show steady improvement for proposed model. Under CF mode, the SCTM-METANET does not show improvement in prediction accuracy, as in the case of predicting TTT. Under FC, the proposed model shows overall improvement. The quantitative measurement of the prediction performance for TTD proves the above observation. During the CC period, the proposed model shows a significant reduction in error (29%), which outperforms the TTT. Under the FC mode, the proposed model shows 5% improvement compared to the baseline model, which is still more satisfying than the case of TTT. The details of quantitative performance comparison can be found in Table 3.

5. Conclusion and Remarks

In this paper, a modified second-order traffic status prediction model, SCTM-METANET, is proposed for improving the performance of the regular METANET model during VSL control. In particular, the proposed model helped to counteract traffic status prediction error during the most congested period. This study employed a switching mode SCTM model and integrated it with the METANET model so that under each of the FC, CC, and CF modes the SCTM-METANET had slightly different parameters. In addition, unlike the original METANET, the proposed model made predictions in a piecewise manner. The original METANET was already implemented in the VSL field test, and the predicted traffic variables were stored, so the collected field data were considered as ground truth data. The proposed SCTM-METANET was calibrated and simulated by a computer, using field data from the same period. Then the prediction results of the original METANET and SCTM-METANET models could be compared to the ground truth data. Data refining was applied to the field raw data of regular METANET predicted results. The concepts of SE and UE were introduced for clarity, and by eliminating SE the prediction result of regular METANET could then be compared to the SCTM-METANET model. The results clearly show that,

during transitional modes FC and CF, the proposed model shows unstable performance, and, during the CC mode, the SCTM-METANET model largely reduces the prediction error of TTT and TTD by 14% and 29%, respectively.

This paper also revealed that both METANET and SCTM-METANET underestimate the traffic density especially within 5pm~6pm when the traffic mode switches to congestion. The happening of this phenomenon is mainly due to the nature of loop detector system and the basic feature of CTM which is the core part of density prediction in both METANET and SCTM-METANET: the road is segmented as cells and it is assumed that within one cell the speed, density, and flow are all homogeneous, and the homogeneous condition transfers to the next cell. Based on this assumption, in the field the freeway is divided into cells which are approximately 500 m long and the loop detectors installed at the beginning of the cell collects and represents the traffic condition of the whole cell, the data collection interval is 20s. Under noncongested situation, the above hardware settings are abundant for accurately estimating traffic conditions. However, when during peak hour, traffic density, speed, and flow are not homogeneous in 500 m and many vehicles would enter the cell in the 20s data collection interval. It is delightful to find that when upgrading METANET to SCTM-METANET, the underestimation is relaxed although not totally eliminated. That is because of the bringing in of supply and demand stochasticity: the model realizes that there exist more vehicles in the cell than previously estimated.

Future research may build on the results of this study by solving the model's problem of instability during the FC and CF modes. Since the current SCTM-METANET model was calibrated offline and with precalibrated parameters, future work with the SCTM-METANET model will involve obtaining parameters and conducting calibration in real time. In the future research, the utilizing of higher frequency data collection method and more location flexible data collection method such as connected vehicle will shed light upon this situation.

Data Availability

The data used in this research is collected by underpavement loop detectors on a segment of urban freeway called Whitemud Drive in Edmonton, Alberta, Canada. The loop detector data collected in the field is continuously sent to the traffic management center of City of Edmonton. The owner of the data is City of Edmonton, the local government. The data is shared by University of Alberta for research use. In this paper, the loop detector data as well as the posted advisory driving speed data was provided by City of Edmonton. To replicate the research results in this paper, one can ask the author for the raw data set but under the approval of City of Edmonton. And recently, all loop detector data of Edmonton can be visualized and downloaded from a public website, although the highest data frequency raw data cannot be downloaded. Using lower frequency loop detector data will partially reproduce the analysis result in this research. The website mentioned is <http://www.its.ualberta.ca/wcpa/homeIndex/index>.

Disclosure

The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the City of Edmonton and Transport Canada. This paper does not constitute a standard specification or regulation.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank Wai Cheung, Janis Chow, Michael Vaudan, Adrian Loh, and Rahim Karmali from the traffic operation group in the City of Edmonton for providing the Whitemud Drive loop detector data. This research work was jointly supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, City of Edmonton, and Transport Canada.

References

- [1] M. Hadiuzzaman, T. Z. Qiu, and X.-Y. Lu, "Variable speed limit control design for relieving congestion caused by active bottlenecks," *Journal of Transportation Engineering*, vol. 139, no. 4, pp. 358–370, 2013.
- [2] Y. Bie, T. Z. Qiu, C. Zhang, and C. Zhang, "Introducing Weather Factor Modelling into Macro Traffic State Prediction," *Journal of Advanced Transportation*, vol. 2017, Article ID 4879170, 15 pages, 2017.
- [3] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.
- [4] C. Antoniou, M. Ben-Akiva, and H. Koutsopoulos, "Online Calibration of Traffic Prediction Models," *Transportation Research Record: Journal of the Transportation Research Board*, pp. 235–245, 2005.
- [5] S. Afandizadeh and J. Kianfar, "A Hybrid Neuro-Genetic Approach to Short-Term Traffic Volume Prediction," *International Journal of Civil Engineering*, vol. 7, p. 41, 2009.
- [6] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.
- [7] A. Spiliopoulou, M. Kontorinaki, M. Papageorgiou, and P. Kopelias, "Macroscopic traffic flow model validation at congested freeway off-ramp areas," *Transportation Research Part C: Emerging Technologies*, vol. 41, pp. 18–29, 2014.
- [8] M. J. Lighthill and G. B. Whitham, "On kinematic waves. II. A theory of traffic flow on long crowded roads," *Proceedings of the Royal Society A Mathematical, Physical and Engineering Sciences*, vol. 229, pp. 317–345, 1955.
- [9] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.
- [10] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic

- theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.
- [11] C. F. Daganzo, "A behavioral theory of multi-lane traffic flow. Part I: long homogeneous freeway sections," *Transportation Research Part B: Methodological*, vol. 36, no. 2, pp. 131–158, 2002.
- [12] C. F. Daganzo, "A behavioral theory of multi-lane traffic flow. Part II: merges and the onset of congestion," *Transportation Research Part B: Methodological*, vol. 36, no. 2, pp. 159–169, 2002.
- [13] H. J. Payne, "Models of freeway traffic and control," in *Mathematical Models of Public Systems: Simulation Council Proceedings Series*, pp. 51–61, 1971.
- [14] J. A. Laval and C. F. Daganzo, "Lane-changing in traffic streams," *Transportation Research Part B: Methodological*, vol. 40, no. 3, pp. 251–264, 2006.
- [15] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141–167, 2005.
- [16] L. Mihaylova, R. Boel, and A. Hegyi, "Freeway traffic estimation within particle filtering framework," *Automatica*, vol. 43, no. 2, pp. 290–300, 2007.
- [17] M. Papageorgiou, J.-M. Blosseville, and H. Hadj-Salem, "Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris," *Transportation Research Part B: Methodological*, vol. 23, no. 1, pp. 29–47, 1989.
- [18] M. Papageorgiou, H. Hadj-Salem, J. M. Blosseville, and N. Bhourri, *Control, Computers, Communications in Transportation*, Elsevier, 1990.
- [19] A. Sumalee, R. X. Zhong, T. L. Pan, and W. Y. Szeto, "Stochastic cell transmission model (SCTM): a stochastic dynamic traffic model for traffic state surveillance and assignment," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 507–533, 2011.
- [20] X. Sun, L. Munoz, and R. Horowitz, "Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control," in *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH 37475)*, vol. 6, pp. 6333–6338, 2003.
- [21] R. Boel and L. Mihaylova, "A compositional stochastic model for real time freeway traffic simulation," *Transportation Research Part B: Methodological*, vol. 40, no. 4, pp. 319–334, 2006.
- [22] A. Sumalee, T. Pan, R. Zhong, N. Uno, and N. Indra-Payoong, "Dynamic stochastic journey time estimation and reliability analysis using stochastic cell transmission model: Algorithm and case studies," *Transportation Research Part C: Emerging Technologies*, vol. 35, pp. 263–285, 2013.
- [23] A. Sumalee, K. Uchida, and W. H. K. Lam, "Stochastic multi-modal transport network under demand uncertainties and adverse weather condition," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 2, pp. 338–350, 2011.
- [24] A. Hegyi, B. De Schutter, and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed limits," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 185–209, 2005.
- [25] Y. Bie, *Implementing Model Predictive Control based Variable Speed Limit on Urban Freeways: Data Imputation, Model Modification and Field Test Analysis*, 2016.
- [26] A. Kotsialos, M. Papageorgiou, and F. Middelham, "Optimal coordinated ramp metering with advanced motorway optimal control," *Transportation Research Record*, no. 1748, pp. 55–65, 2001.
- [27] L. Elefteriadou, S. S. Washburn, Y. Yin, V. Modi, and C. Letter, *Variable Speed Limit (VSL) – Best Management Practice*, vol. 11, Florida Dep Transp Res Center, 2012.
- [28] X. Wang, M. Seraj, Y. Bie, T. Z. Qiu, and L. Niu, "Implementation of variable speed limits: Preliminary test on Whitemud Drive, Edmonton, Canada," *Journal of Transportation Engineering*, vol. 142, no. 12, 2016.

Research Article

A Two-Layer Network Dynamic Congestion Pricing Based on Macroscopic Fundamental Diagram

Bangyang Wei ^{1,2} and Daniel(Jian) Sun ^{1,2}

¹State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai, China

²Center for UAV Applications and ITS Research, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China

Correspondence should be addressed to Daniel(Jian) Sun; danielsun@sjtu.edu.cn

Received 22 February 2018; Revised 15 May 2018; Accepted 27 May 2018; Published 1 August 2018

Academic Editor: David Z. W. Wang

Copyright © 2018 Bangyang Wei and Daniel(Jian) Sun. 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.

Dynamic congestion pricing has attracted increasing attentions during the recent years. Nevertheless, limited research has been conducted to address the dynamic tolling scheme at the network level, such as to cooperatively manage two alternative networks with heterogeneous properties, e.g., the two-layer network consisting of both expressway and arterial network in the urban areas. Recently, the macroscopic fundamental diagram (MFD) developed by both field experiments and simulation tests illustrates a unimodal low-scatter relationship between the mean flow and density network widely, providing the network traffic state is roughly homogeneous. It reveals traffic flow properties at an aggregated level and sheds light on dynamic traffic management of a large network. This paper proposes a bilevel programming toll model, incorporating MFD to solve the unbalanced flow distribution problem within the two-layer transportation networks. The upper level model aims at minimizing the total travel time, while the lower level focuses on the MFD-based traffic assignment, which extends the link-based traffic assignment to network wide level. Genetic algorithm (GA) and the method of successive average were adopted for solving the proposed model, on which an online experimental platform was established using VISSIM, MATLAB, and Visual Studio software packages. The results of numerical studies demonstrate that the total travel time is decreased by imposing the dynamic toll, while the total travel time savings significantly outweigh the toll paid. Consequently, the proposed dynamic toll scheme is believed to be effective from both traffic and economic points of view.

1. Introduction

Congestion pricing has been regarded as an effective traffic demand management policy that has been applied in many cities, among which Singapore was the first one to impose road congestion pricing at city wide level [1]. The development process has experienced two stages from the initial regional license scheme to the real-time dynamic electronic charging system. In United States, most tolled roads are related to new lanes or lanes that have been opened to High Occupancy Vehicles (HOV). In California, private investors have built the 91st fast lane based on the existing state highway S91, providing drivers the option of using an uncharged lane or a time-varying fee lane. These successful road congestion pricing schemes have aroused wide considerations in exerting

more efficient and impartial pricing strategies in large urban networks.

The theory of congestion charging was developed by Walters [2] and Vickrey [3] that the users should pay for both the toll equal to their own travel cost and the additional congestion cost imposed on other users due to the extra travel occupying the public road resource, which is consistent with the concept called “marginal cost” in economics. Sheffi [4] found that the additional cost added in the static models can drive the user equilibrium pattern to achieve system optimum. Ferrari [5] studied the charging problem for urban transport network with elastic demand and link capacity constraints and found that the model has no solution which satisfies the capacity constraints, but additional costs can be imposed on some links so that an equilibrium solution

that satisfies the capacity constraints always exists. In terms of multiuser classes, Leurent [6] designed a model that allows for differentiating travelers by means of an attribute called value of time that is continuously distributed over a real interval. Yang and Lam [7] and Yang and Bell [8] studied congestion pricing in the queue network under fixed and elastic demand, respectively, through establishing the bilevel programming model, where the lower level problem describes users' routes choice behavior under condition of both queuing and congestion, while the upper level problem is to determine road toll to optimize a desired system's performance. In view of the time-varying traffic demand, the static toll models need to be extended to dynamic toll models. Vickrey [9] applied deterministic queuing theory to firstly propose the dynamic pricing scheme called the bottleneck model that the travelers should pay the toll for eliminating the queuing delay without toll. Yang and Meng [10] combined the application of the space-time expanded network representing the time-varying traffic flow and the conventional network equilibrium modeling technologies. Carey and Watling [11] extended the formulation that used Cell Transmission Model (CTM) for constructing system optimum (SO) formulation to allow more general nonlinear flow-density functions and found that if the tolls computed from the dynamic system optimum (DSO) solution are imposed on the users, the DSO solution would also satisfy the criteria for a dynamic user equilibrium (DUE). Ban and Liu [12] formulated a link-nodes discrete-time dynamic toll model as a bilevel problem, where the upper level is to minimize the total weighted travel time, while the lower level is to capture the users' route choice behavior. Laval and Castrillón [13] analyzed the time-dependent congestion pricing scheme on two alternative routes that have different bottleneck capacities and surprisingly found that the minimum total system delay can be achieved with many different pricing strategies, which are valuable for traffic management. Although both the static and dynamic settings and cases that have been addressed so far are encouraging, there are some certain deficiencies: (i) traffic system is not in steady state at link level validated by simulation evaluation and field experiment, and thus toll calculated based on ideal curve is not optimal [14, 15]; (ii) existing studies on dynamic congestion pricing have been limited to certain idealized networks, because the model and computation of the link-based dynamic traffic assignment (DTA) are so complex that these research cannot be applied to a large city network practically.

Recent findings on the traffic dynamics at the network level named macroscopic fundamental diagram (MFD) have aroused wide attentions. Geroliminis and Daganzo [14, 15] used both detector data from microscopic simulation test for the San Francisco Business district in California and a data combination of fixed detectors and floating vehicle probes as sensors from field experiment in downtown Yokohama in Japan to find that MFD relating the number of vehicles in network to space mean flow or outflow presents a unimodal low-scatter relationship, if the road network is in homogenous congestion state. Scholars mainly focus on the research about its nature, influence factors, and application. Leclercq and Geroliminis [16] took advantage of the topology of the route

and the variation theory to obtain MFD, in order to overcome the homogeneity rule which points out the congestion state at link level should keep pace. The influencing factors on the shape of MFD have also been investigated gradually. Geroliminis and Sun [17] found the "hysteresis phenomenon" that a clockwise loop exists between the flow and density instead of linear curve and then provided the explanation with density inhomogeneity and lack of adequate data. In terms of the influence of turning flow in intersections on MFD, Geroliminis et al. [18] found that the left traffic flow reduces the largest value of MFD. Because of the increasing recognition about MFD, researchers have started to focus on how to apply MFD to region traffic management. One of the main tasks was to investigate network division, thus obtaining a well-defined MFD. The most classic method was developed by Ji and Geroliminis [19] that divided the entire network according to the congestion feature [20, 21], and then the dynamic division problem was also studied. Keyvan-Ekbatani et al. [22] studied the feedback gate control method using the simulation network with perimeter gate control and obtained satisfying results with lower total travel time. Aboudolas and Geroliminis [23] used multireservoir networks with well-defined MFDs to design the perimeter and boundary flow control schemes that aimed at distributing the accumulation of vehicles in each reservoir as homogeneously as possible. However, deficiencies are also existed in these control schemes: (i) the change of signal at the cordon may have an influence on the shape of MFD, but the influence was largely ignored; (ii) the common perimeter control needs sufficient road space for stopping and queuing [24, 25], which may not be always allowed. Cordon-based or area-based dynamic congestion pricing at network level has been limited to some simple network due to the complexity of link-based dynamic assignment. Geroliminis and Levinson [26] combined the Vickrey's bottleneck theory with MFD to propose a cordon-based congestion scheme, which is easier implemented in real road networks because of more convenient data collection and much easier computation. Zheng et al. [27] adopted an agent-based simulation to develop and apply a cordon-based dynamic pricing scheme, in which tolls are controlled by MFD. The above works shed light on how to apply MFD to design the dynamic congestion pricing scheme at network level.

This paper aims to combine the MFD with dynamic traffic assignment theory to design a dynamic pricing scheme in heterogeneous networks. A bilevel optimal equilibrium model combining the MFD theory is proposed, which is consistent with the traffic dynamics at network level and can also timely carry out online data analysis and output the expected toll, due to fewer data collection and lower computation cost than link-based equilibrium model. The upper level problem is to minimize the total travel time. The lower level problem is a network equilibrium model with the MFD. In detail, the upper level model is to obtain the optimal toll for the designed system objective from the manager perspective, while the lower level model is to output the expected flow distribution by solving the dynamic user equilibrium assignment with incorporating MFD. Genetic algorithm and method of successive average were used

together to solve the bilevel model. Finally, a numerical test for the two-layer network consisting of the loop expressway network and the linear arterial streets was employed in the simulation environment, which combines the microscopic simulation software VISSIM, mathematic solver software MATLAB, and the development tool Visual Studio as the medium to implement the proposed bilevel toll model.

The remainder of the paper is organized as follows. Section 2 introduces the important features of MFD and explains the reliability for applying MFD to dynamic congestion pricing scheme on the two-layer network. Section 3 describes a bilevel dynamic toll programming model, the solving algorithm, and scheme for model application. In Section 4, the proposed model is tested with the two-layer network in the simulation-feedback environment. Finally, conclusions and recommendations for future work are provided in Section 5.

2. Features of MFD and Research Question

2.1. Features of MFD. Technically, MFD represents traffic characteristics at the network level by aggregating the link flow and density [14, 15]. The weighed space mean density K and the weighed space mean flow Q for a road network with the homogeneous traffic distribution can be expressed as follows:

$$K = \frac{\sum_{\Psi} k_{\Psi} \cdot l_{\Psi} \cdot n_{\Psi}}{\sum_{\Psi} l_{\Psi} \cdot n_{\Psi}} = \frac{N}{L} \quad (1)$$

$$Q = \frac{\sum_{\Psi} q_{\Psi} \cdot l_{\Psi} \cdot n_{\Psi}}{\sum_{\Psi} l_{\Psi} \cdot n_{\Psi}} = \frac{P}{L} \quad (2)$$

where $\Psi \in \Phi$ denotes an individual link in the network links set; k_{Ψ} is the traffic density of link Ψ ; l_{Ψ} denotes the length of link Ψ ; n_{Ψ} is the number of lanes of link Ψ ; q_{Ψ} is the traffic flow of link Ψ ; N called the accumulation is the existing number of vehicles within the current network; P is the network travel production; and L denotes the total length (lane-kilometers) of the network.

The network space mean speed V can be expressed as follows:

$$V = \frac{Q}{K} \quad (3)$$

A representative fundamental diagram relating the space mean flow to space mean density is resemble with the Figure 1.

MFD can be divided into three typical regimes: “free flow” region A with fewer vehicles within the network, which have relative free speed; “maximum flow” region B, where the space mean flow reaches the highest point with the related optimal mean density value called “critical density” that many researchers take advantage of the value to study network control strategies; “congestion flow” region C, where the space mean density is beyond the critical density, the space mean flow decreases with the mean density increasing, and then the congestion will occur. When network is in state D, no one car can nearly arrive to the destination.

Regarding the road network as a reservoir, the state formula of the network system is described by the mass conservation equation as below:

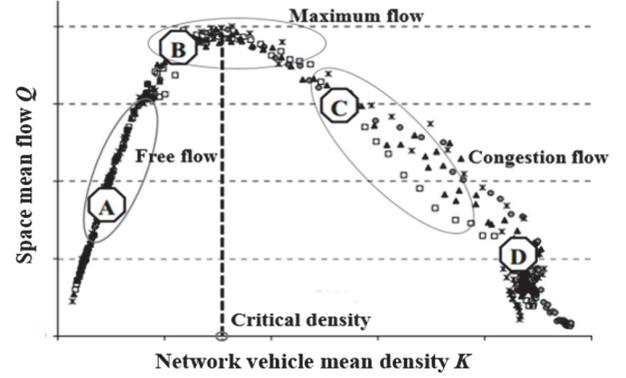


FIGURE 1: Representative MFD (space mean flow versus mean density) (source: reference [14]).

$$\frac{dN(t)}{dt} = q_{in}(t) - q_{out}(K(t)) \quad (4)$$

where $q_{in}(t)$ is the inflow to the network at time t and $q_{out}(K(t))$ is the total outflow representing the trip completion rate at time t . To further analyze the outflow, there is also the unimodal relationship with the space mean density and a robust liner relationship with the space mean flow, which is regarded as another important finding according to field experiment in downtown Yokohama [15].

These formulas provide two important clues: (i) the link-based state formula can be extended to region-based equation thanks to the $q_{out}(t)$ which can be predicted by the state variable $K(t)$; (ii) different from the bottle model that queue will occur when demand goes beyond capacity, the state equation illustrates that traffic system are dynamic, and the average travel time is connected with the initial state of the system as well as input flow.

The above description provides preliminary recognition about the concept of MFD, and features are summarized as follows: (i) MFD is the characteristic of network and not very sensitive to different demand patterns; (ii) the space mean flow can maintain the maximum value when the space mean density value is adjacent to the critical density value; (iii) using existing loop detector data and GPS data can obtain the MFD accurately. Considering these features of MFD, it can be applied to dynamic network cordon pricing.

2.2. Problem Description. Urban transportation network generally can be divided into multiple layers with different properties, such as arterial street, expressway, and metro network [28]. Expressway is a type of road that is built for fast travel with higher design speed [29] and can only be accessed by ramps. Arterial streets are high-capacity urban roads and serve to carry long-distance flows between important centers of activity, which produce large amounts of carbon emission [30]. Compared to the arterial street, expressway has the continuous traffic flow that vehicles do not have to stop for the signal lights in intersections. The expressway is commonly built above the land to decrease land usage and avoid grade crossing in city center region, however, the arterial street is often built on the land for more flexible access. The network

system consisting of the expressway and arterial streets is called the two-layer network.

Drivers often have dilemmas when choosing between the arterial road network and the loop expressway, and the phenomenon exists sometimes that one is congested but the other is in a low density. Dynamic congestion pricing has been proved as an effective strategy to make the users change route by changing the trip utility for the optimal demand distribution on the network. Therefore, there is practical need to develop toll strategy at regional level by means of macroscopic traffic flow theory which is consistent with the network traffic dynamics.

3. Model Formulation

Urban traffic network MFD provides a unimodal low-scatter relationship between space mean density and space mean flow or outflow, if link flows are in roughly homogeneous congestion state within the region. A smoothly decline curve between network mean speed and density can also be identified. The urban network consisting of the multilayers with different properties can be represented by the multi-MFDs respectively. Considering the practical problem, it is essential to take proper toll strategy for balancing the traffic demand in two-layer network consisting of the loop expressway network and arterial streets. The following sections introduce the bilevel model combining MFDs and traffic assignment theory to design a dynamic toll scheme.

3.1. The Lower Level Model. The lower level model combines DUE model with MFD to develop an equilibrium assignment model at aggregated level which regards the network as a special reservoir. Comparing to the traditional link-based dynamic traffic assignment model, the network-based user equilibrium model has some modifications: (i) the trip completion rate is polynomial relationship with the space mean density in network; (ii) the travel time in network depends on the average travel distance and the space mean speed, which can be calculated through the formulation of space mean flow divided by space mean density, and the mean speed is approximately linear relationship with the mean density.

Travel costs include travel time consuming, fuel cost, and road toll. This paper adopts generalized transportation cost function that all travel costs are calculated as equivalent hours using fixed value of time (VOT). Travel costs including travel time and toll are considered as follows:

$$\hat{t}_a(K_a(t), u_a(t)) = t_a(K_a(t)) + \frac{u_a(t)}{VOT} \quad (5)$$

$$t_a(K_a(t)) = \frac{l_a}{V_a(K_a(t))} \quad (6)$$

where $\hat{t}_a(K_a(t), u_a(t))$ is the total travel cost in network a at time t , $u_a(t)$ is the toll in network a at time t , $t_a(K_a(t))$ is the planned travel time in network a at time t ; $K_a(t)$ is space mean density of the network a at time t ; l_a is the average travel distance in network a , $V_a(K_a(t))$ is space mean speed in the network a at time t .

Similar to the formula of link-based dynamic user equilibrium, the objective function is to minimize the integral function representing travel within the two reservoir systems in a time interval T . The MFD-based dynamic user equilibrium model can be formulated as follows:

$$J_l = \min \sum_{a \in A} \int_{i \cdot T}^{(i+1) \cdot T} \int_{K_a(i \cdot T)}^{K_a(t)} \hat{t}_a(w, u_a(t)) dw dt \quad (7)$$

subject to

$$\frac{dN_a(t)}{dt} = \frac{dK_a(t) \cdot L_a}{dt} = q_{in}^a(t) - q_{out}^a(t), \quad \forall a, t \quad (8)$$

$$\sum_{a \in A} q_{in}^a(t) = q_{in}(t), \quad \forall a, t \quad (9)$$

$$q_{out}^a(t) = q_{out}^a(K_a(t)), \quad \forall a, t \quad (10)$$

$$q_{in}^a(t) \geq 0,$$

$$q_{out}^a(t) \geq 0,$$

$$k_a(t) \geq 0,$$

$$\forall a, t$$

where (8) is the network state equation for the network $a \in A$; constraint (9) is the node inflow equilibrium equation; L_a is the total lane length for network a ; $q_{in}(t)$ is the overall traffic demand at time t ; and $q_{in}^a(t)$ is the inflow to network a at time t . Equation (10) is the outflow which is the function of space mean density; constraint (11) ensures that all variables satisfy nonnegative condition.

3.2. The Upper Level Model. The upper level problem is to investigate the desired objective from the manager perspective, in which various purposes may be included within the previous literatures, such as minimizing the total travel time [31, 32]. In the simulation experiment, the total travel time can be formulated as the integral of the number of vehicles (vehicle-hours) in the two-layer network as follows:

$$J_{up} = \min \int_{i \cdot T}^{(i+1) \cdot T} \sum_{a \in A} N_a(t) dt \quad (12)$$

$$u_a(t) \leq u_{max} \quad (13)$$

where $N_a(t)$ is the number of vehicles in network a at time t ; u_{max} is the allowed maximum toll.

3.3. Solution Algorithms. A common characteristic of the model is to determine the optimal value in a set of decision variables, in order to keep the network performance in optimal. Typical examples include network design and road pricing. These questions can be depicted as a Stackelberg game or leader-follower game [33]. The leader refers the traffic manager with the goal of optimizing network usage at system level, and the follower refers the user with optimal

path choice based on route utility. Because the upper and lower problems are mutually restricted and the model is non-convex, it is difficult to solve the bilevel programming model by means of the general optimization method. For the lower model, the model is processed discretely at time dimension. The discrete time form can be considered as a nonlinear programming problem, so the Frank-Wolf algorithm or the method of successive average can be applied to solve the function. From the toll perspective, as time-varying pricing is difficult to be implemented, it is reasonable to adopt constant toll during each evaluation interval.

The above discussion provides the idea that the bilevel model is discrete at time dimension during the evaluation time interval T with the fixed toll. Genetic algorithm with the characteristics of global search is adopted to solve the upper level problem. The lower model is discrete at time dimension. The method of successive average called the approximate equilibrium algorithm is applied to traffic assignment when the calculation accuracy is not strict and time saving is needed. At each iteration, the all-or-nothing assignment is conducted to create auxiliary path flows which are weighted together with the path flows from the former iteration, to obtain path flows for the next iteration. Iterations do not terminate until the algorithm satisfies the convergence criteria and the final solution is approximately equal to the equilibrium solution. These user equilibrium assignment algorithms are common in many research, and the attention point is that, for each time segment, the number of vehicles in each network is so connected with the former stage as the transition process. To simplify the calculation, the instantaneous travel time is assumed to be equal to actual travel time. This study combines Genetic Algorithm with the method of successive average to solve the bilevel programming model, as follows:

Step i. Generate initial population and determining cross-over rate, mutation rate of genetic algorithm, the total number of individuals in each generation, and the maximum evolutionary algebra M . Set evolutionary algebra $m = 0$.

Step ii. For each individual, apply the assignment algorithm to solve the lower level model for each discrete interval, and the state equation $N(j+1) = N(j) + (q_{in}(j) - q_{out}(j)) \cdot \Delta t$ is regarded as the transient process. Then, each responding flow distribution among the group is obtained.

Step iii. Calculate the function value of the upper model corresponding to each toll and obtaining the fitness value. Choose the optimal solution. If $m > M$, output the best solution; otherwise, set $m = m + 1$.

Step iv. Cross over, mutate, and select individuals according to the fitness index value and generate new offspring; then go to Step ii.

3.4. Scheme for Model Application. The proposed model presents the dynamic toll problem as a bilevel problem. The upper level is to minimize total travel time and the lower

level is the dynamic user equilibrium at the network level. Managers provide a charging rate through real-time detection and analysis of road network data, then the users adjust travel routes according to timely utility so that the two alternative networks can maintain the satisfying state. The toll strategy takes full use of the timely traffic information and provides the predicted toll for the given time interval T , with the entire simulation time set as T_a . The timely feedback toll strategy including two stages for the MFD-based bilevel toll model is shown as follows:

First Stage. The two-layer network MFDs were obtained respectively, including the diagrams of network mean flow and mean density, mean speed and mean density, trip complementation rate, and mean density.

Then, the following steps are incorporated as the **second stage** of the model application.

Step i. To simulate for several minutes at start, collect state information for networks, and set $i = 0$.

Step ii. To solve the bilevel programming model to obtain the predictive toll for the next interval according to the timely collected data feedback and MFDs obtained from the first stage, $i = i + 1$.

Step iii. To change the network toll timely and the user will choose the route according to the dynamic utility timely. If $i \cdot T < T_a$, deliver the network traffic data to Step ii; otherwise, stop the simulation experiment and evaluate the result.

4. Case Studies and Results

To test the bilevel model and the corresponding algorithm proposed in the above discussion, a simulation experiment of two-layer network was employed. This study integrated VISSIM, MATLAB, and the development tool Visual Studio as the main control program. In detail, VISSIM and Visual Studio were integrated through COM data communication interface, and Visual Studio and MATLAB were integrated through Dynamic Link Library (DLL). VISSIM was used to simulate traffic flow, MATLAB was used to solve the bilevel programming model, and Visual Studio was used as the intermediate medium.

4.1. MFD for Simulation Network. There are many cities using the circular expressway network to relieve the pressure of transit traffic in the center business region. Paris in France and Chengdu in China are two typical instances as shown in Figures 2(a) and 2(b), which can be summarized as Thomson's urban structure model [34] as shown in Figure 2(c), consisting of grid and ring network. It is suitable to the city center with developed economy and star towns around the perimeter. The simulated two-layer network consists of a loop expressway network and linear arterial streets as shown in Figure 2(d). In the simulation network, the expressway network refers to the bidirectional loop roads including 32 links with two lanes around 500m in length and 3.5m in

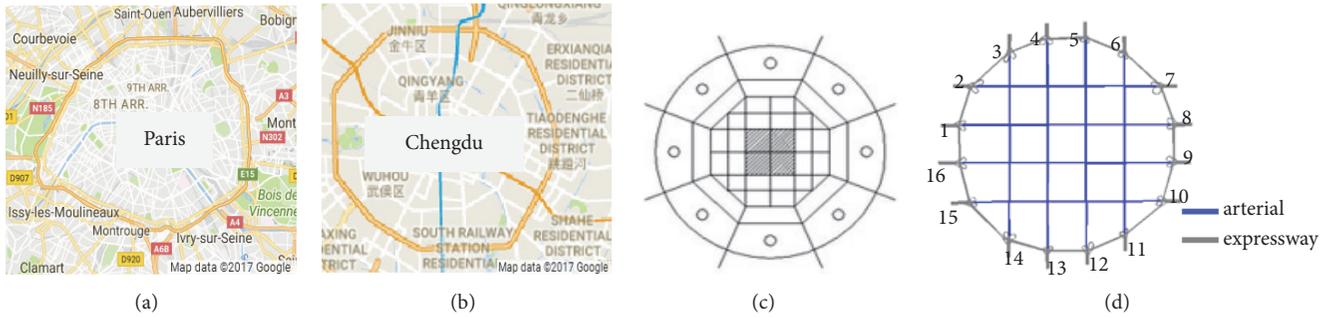


FIGURE 2: Two-layer network layout: (a) city center for Paris, France, (b) city center for Chengdu, China, (c) Thomson urban structure model, and (d) simulation two-layer networks layout.

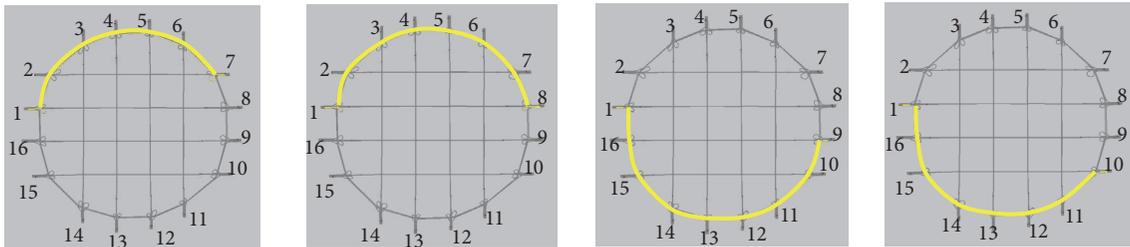


FIGURE 3: Routes in expressway (e.g., Origin 1 with Destinations 7, 8, 9, and 10).

width. The arterial streets network refers to the bidirectional diameter roads including 80 links with two lanes around 500m in length and 3.5m in width. The intersections of arterial streets network adopt pretimed fixed timings, with both Green and Red set as 30s. The free flow speeds within the links of expressway and arterial road are set as 60km/h and 35km/h, respectively.

16 traffic zones are located at the perimeter with the origins and destinations uniformly distributed across the network. To avoid the large variance of trip distance along each network, the OD pairs with the neighbor nodes are not considered. For example, the simulation experiment only considers Origin 1 with Destinations 7, 8, 9, and 10 and Origin 2 with Destinations 8, 9, 10, and 11. As a result, the number of all OD pairs is 64.

Considering the objective is to investigate the influence of the dynamic toll on the route choice between the expressway and arterial streets, so as to achieve a more efficient network performance, e.g., minimizing total travel time, the traffic assigned on routes in each network is simply considered as static instead of dynamic. Many research papers adopted microsimulation that used fixed routing scheme obtained by the static traffic assignment (STA) method [35] to obtain MFD. For example, CORSIM was used in San Francisco, California [14], and VISSIM was used in Orlando, Florida [36], and in Zurich, Switzerland [37]. Recent studies have adopted the dynamic traffic assignment (DTA) modules in VISSIM to obtain the MFD [38]. In dynamic network loading, DTA simulation method is iterative process to search new routes according to the experienced travel time that attempts to provide users equilibrium traffic states. However, DTA method may be not appropriate in this research, which

could induce large variance for the time-varying average trip length in each network due to the fact that some drivers would choose the long detour to avoid the congestion [39]. Moreover, DTA method is more computationally demanding due to a few iterative processes and more unstable as convergence is not always guaranteed. As the research question is to investigate the applicability of dynamic toll at the large network level using the MFD rather than link-based flow-density function, the research simply adopts the STA method with capacity restraint to obtain the fixed routes with the initial traffic demand pattern and then keeps the same OD pattern and multiplies the factors to scale the OD demand for achieving an integrated MFD [37, 40]. The resulting time along each layer is related to the travel distance and network density within the simulation experiment. The average travel distances and the standard deviation (SD) for expressway and arterial streets are shown in Tables 1 and 2, and the typical routes in expressway and arterial streets for Origin 1 with Destinations 7 to 10 are shown in Figures 3 and 4.

The routes of all OD pairs in expressway can be categorized into three types (average trip distance: 3500m, SD: 354m), and the routes of all OD pairs in arterial streets can be categorized into four types (average trip distance: 3030m, SD: 363m). The average trip distance is expected to be constant that the average trip distance is 3.5km for expressway network and 3.0km for arterial streets.

The initial experiment is to obtain the MFDs of the two-layer alternative networks, respectively. The origin-destination (OD) matrix is used to express traffic demand with uniform distribution. The demand factor is used to modify the demand amount for the intact MFD. Detectors are installed at every 100m length in each link and the related data

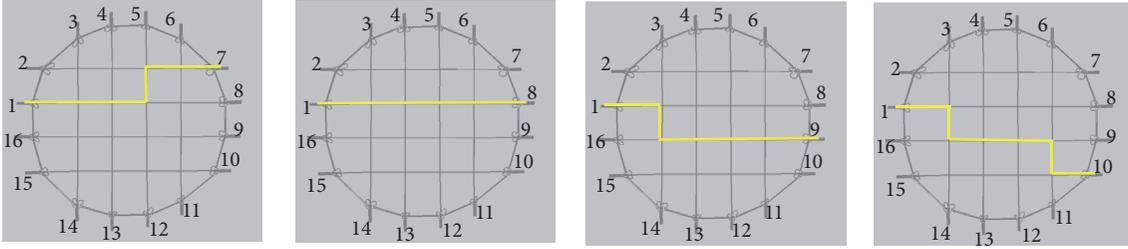


FIGURE 4: Routes in arterial streets (e.g., Origin 1 with Destinations 7, 8, 9, and 10).

TABLE 1: Average trip distance for expressway.

Trip distance (m)	3000	3500	4000
Number of OD pairs	16	32	16
Proportion	25%	50%	25%
Average travel distance (m)	3500 SD = 354		

TABLE 2: Average trip distance for arterial streets.

Trip distance (m)	2500	2750	3250	3500
Number of OD pairs	8	24	16	16
Proportion	12.5%	37.5%	25%	25%
Average travel distance (m)	3030 SD = 363			

including flow and occupancy are collected and aggregated at every 120 seconds. The results are shown in Figure 5.

By adopting least square fitting, six polynomial functions were obtained. For expressway network, space mean flow versus mean density can be expressed as the formula: $Q = 0.005K^3 - 1.337K^2 + 88.923K (R^2 = 0.96)$, space mean speed versus mean density can be expressed as the formula: $V = -0.98K + 84.92 (R^2 = 0.97)$, and network outflow versus mean density can be expressed as the formula: $q_{out} = 0.059K^3 - 13.35K^2 + 751.76K (R^2 = 0.97)$. For arterial network, space mean flow versus mean density can be expressed as the formula: $Q = 0.0062K^3 - 1.15K^2 + 49.76K (R^2 = 0.98)$, space mean speed versus mean density can be expressed as the formula: $V = -0.51K + 36 (R^2 = 0.90)$, network outflow versus mean density can be expressed as the formula: $q_{out} = 0.027K^3 - 11.25K^2 + 634K (R^2 = 0.91)$.

4.2. Dynamic Pricing Experiment Evaluation. As indicated in Figures 5(a), 5(b), 5(d), and 5(e), the critical densities for expressway network and arterial streets are both about 40veh/km, and the network mean speeds for expressway network and arterial streets at the critical density are around 40km/h and 20km/h, respectively. According to user equilibrium principle, due to higher mean speed for expressway network even though in right band of MFD which is larger than the critical density, when in high traffic demand condition, the expressway network would operate in congestion flow state which is easy to cause congestion; however, the arterial streets would operate in the free flow state. The purpose of the simulation experiment is to decrease the density of expressway network to maintain the two-layer density load

TABLE 3: Improvement with applying the toll scheme.

Index	No toll	Toll	Improvement
Total Travel Time (h)	6415.6	5080.8	1334.8
Total Collected Revenue (\$)	-	13620.2	13620.2
Average Speed (km/h)	33.2	42.3	9.1
Average Travel Cost (h)	0.123	0.114	0.009

in an optimal state. Consequently, only the freeway network is charged with extra toll in the simulation experiment considering the significant difference of average speeds between the two-layer networks, however, if in field experiment, the two-layer networks would cooperatively take toll schemes. In this research, deterministic and homogeneous population characteristics are assumed with fixed VOT 16\$/h [41, 42], and the maximum toll is set as 10\$ [43].

The fluctuated demand profile including two peak periods is presented in Figure 6.

In this section, simulation experiments for applying both the dynamic toll strategy and no toll were carried out. The simulation for first 10 minutes is to produce preliminary volume in the network, and the data about network state are collected and aggregated for each 2 minutes. Table 3 presents the improvements after applying the toll scheme.

The total travel time savings and the total collected revenue were investigated. The total travel time decreases from 6415.6h to 5080.8h. The total collected revenue is 13620.2\$. Considering the value of time is 16\$/h, the total toll paid in the study network is smaller than the total time savings, by 36.2%. The time savings are much higher than the toll paid, which is a promising result at an aggregated level. The average travel cost without toll is 0.123h, and the average travel cost with toll represented by time cost decreased to 0.114h. Moreover, the average speed increased by 9.1km/h.

The time-dependent tolls are presented in Figure 7, while the average speeds with toll or not were compared in Figure 8. It is found that the time-dependent toll diagram also has two peaks, and the highest toll values are 0.66\$ per trip and 0.69\$ per trip in the fourth and ninth periods, respectively. For the two high peak demand periods, the average speeds with no toll are only 29.5 km/h and 26.1 km/h, while the average speeds with toll are 39 km/h and 38.3 km/h. These further illustrate that the toll scheme is effective in improving the speed for the two-layer network. Moreover, although tolls at all time are above some base level of about 0.45\$, under

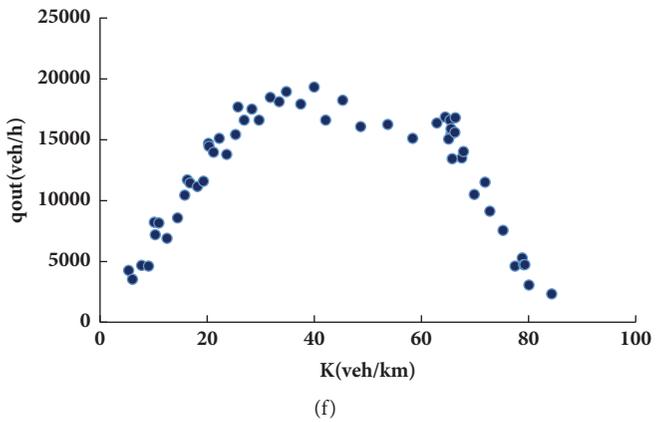
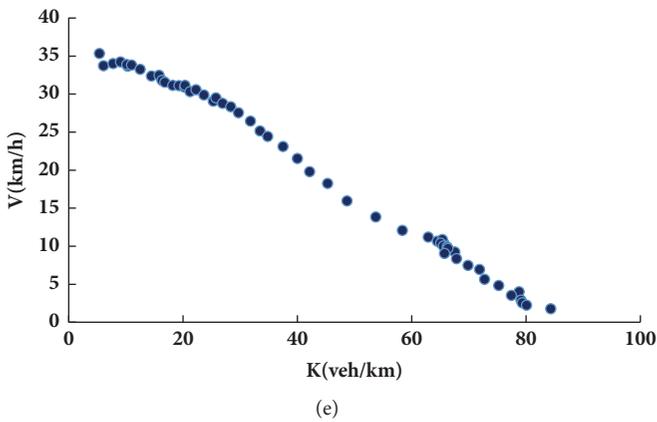
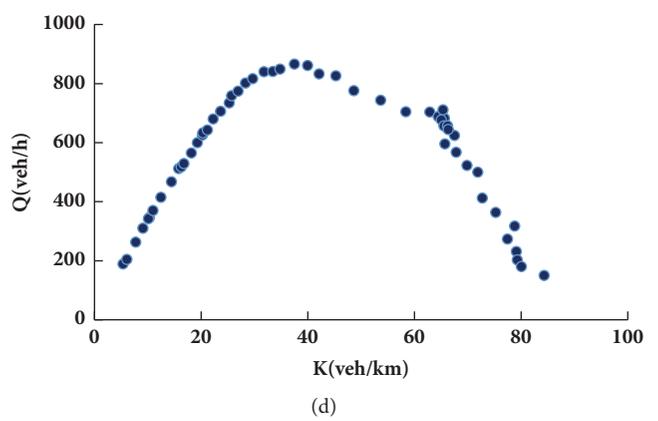
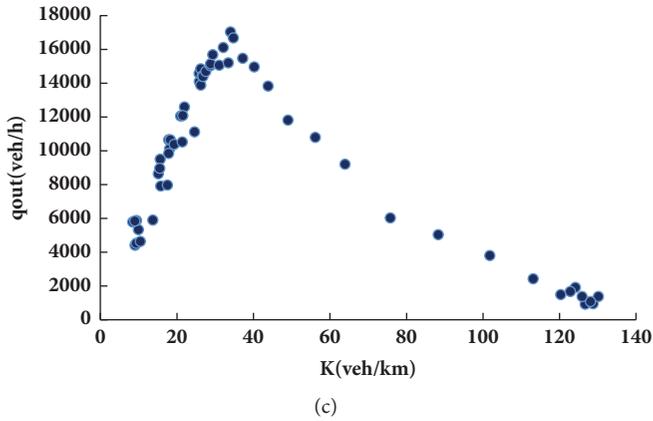
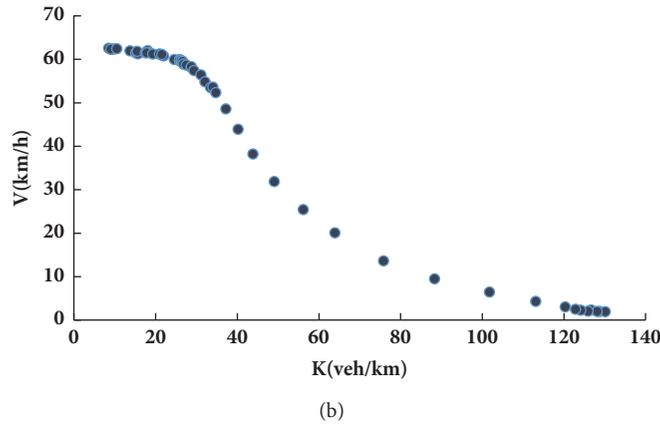
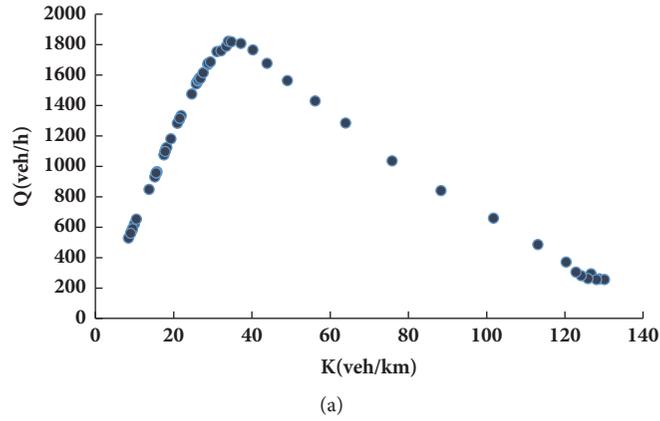


FIGURE 5: MFD of the two-layer network: (a) expressway network: space mean flow versus mean density, (b) expressway network: space mean speed versus mean density, (c) expressway network: network outflow versus mean density, (d) arterial streets: space mean flow versus mean density, (e) arterial streets: space mean speed versus mean density, and (f) arterial streets: network outflow versus mean density.

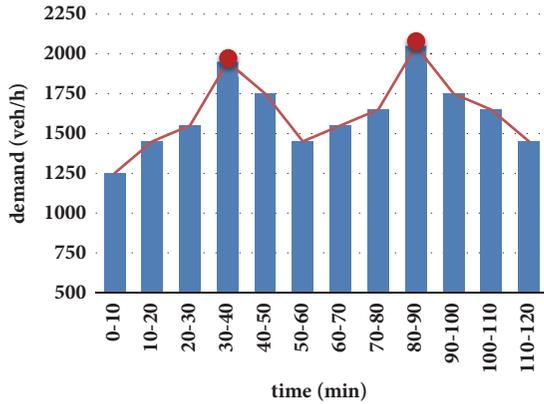


FIGURE 6: Fluctuated demand profile including two peak periods.

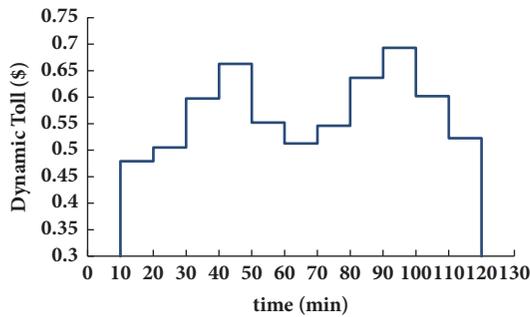


FIGURE 7: Time-dependent toll.

bottleneck models and departure time choice, the criterion value may tend to be null as the base level.

Furthermore, Figures 9(a) and 9(b) present the detailed MFD for expressway and arterial streets in toll and no toll conditions. Under no toll condition, more drivers would choose expressway for the higher travel speed, then the average density is far beyond the critical density (Figure 9(a)), which is much easier to cause traffic congestion. However, the average density of expressway falls on the left part in toll. For arterial streets, the range of the average density is from 5 to 35veh/km, while it mainly falls between 15 to 25veh/km under toll condition (Figure 9(b)), which remains a larger reservoir capacity for more traffic demand. It can be explained that during toll condition the road networks operate in much more optimal state by analyzing the MFD.

5. Conclusions

In this paper, a dynamic toll strategy at a two-layer network is proposed. The idea is to adopt a macroscopic traffic model, i.e., the macroscopic fundamental diagram (MFD), to determine the optimal toll. The bilevel programming model combining the traffic assignment principle and the theory of MFD is proposed. The upper level model is to decrease the total travel time in the two-layer network by optimizing the toll and the lower level model extends the link-based dynamic user equilibrium to network-based dynamic user equilibrium by means of the physics of network traffic expressed by

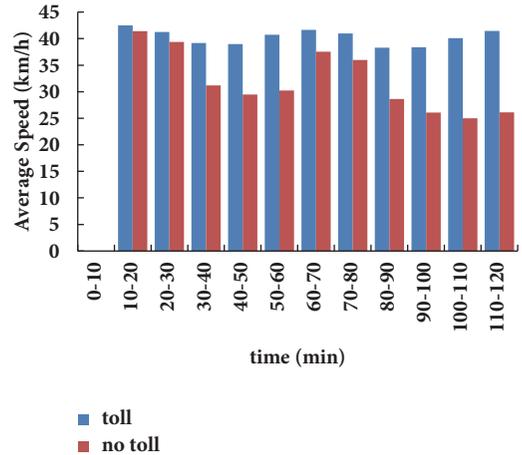


FIGURE 8: Average speed in toll and no toll condition.

the MFD. Genetic algorithm and the method of successive average are applied for the solution to the discretized bilevel model. The integration of VISSIM, MATLAB, and Visual Studio as the main control program is proposed to validate the proposed bilevel toll model for the given two-layer network.

The results indicate that the real-time toll strategy is effective for keeping the network operating in a better state. On other words, the arterial road resources can be utilized in a more effective way and the expressway network becomes more moderately to avoid the road congestions. Compared with no toll condition, in which UE criterion depends just on the average travel time, the bilevel toll model can significantly decrease the total travel time. Because the MFD-based dynamic toll model at aggregated level can overcome the calculation complexity of link-based assignment and the data collection and statistics to evaluate the network state are easier, the proposed dynamic toll strategy is anticipated to become more effective in a larger urban network.

The proposed dynamic tolling is promising to solve the exist problem in charging toll at network aggregated level for the optimal traffic distribution on the overall urban network. However, some problems need more efforts to solve in future. Users with different value of time (VOT) and heterogeneous travel distances should be thought, and the agent-based simulation can also be used to implement the toll model [44–46]. Moreover, the multiple modes traffic should be incorporated considering the toll is also to encourage users to take the public transport. To design more practical management scheme for the multi-MFDs regions, it is especially necessary to develop effective algorithms to decompose the large-scale center network into interconnected subdomains represented by the own MFD, such as the shortest distance decomposition algorithm for partitioning large-scale network [47].

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

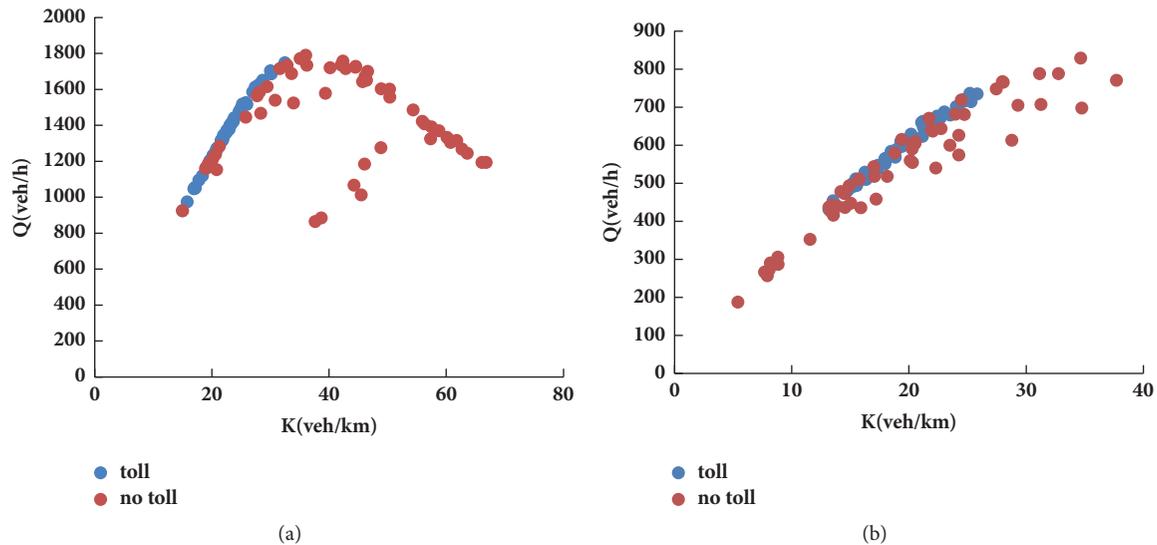


FIGURE 9: MFD in toll and no toll conditions: (a) for expressway; (b) for arterial streets.

Disclosure

Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to express their appreciation to Dr. Zhong-Ren Peng from Center for UAV Applications and ITS Research, School of Naval Architecture Ocean and Civil Engineering, Shanghai Jiao Tong University, for his valuable suggestions and assistance in this study. The research was supported in part by the Major Project of National Social Science Foundation of China [16ZDA048], the Humanities and Social Science Research Project, Ministry of Education, China [18YJCZH011], and the Shanghai Municipal Natural Science Foundation [17ZR1445500], China.

References

- [1] S.-Y. Phang and R. S. Toh, "From manual to electronic road congestion pricing: The Singapore experience and experiment," *Transportation Research Part E: Logistics and Transportation Review*, vol. 33, no. 2, pp. 97–106, 1997.
- [2] A. A. Walters, "The theory and measurement of private and social cost of highway congestion," *Econometrica*, vol. 29, no. 4, pp. 676–699, 1961.
- [3] W. Vickrey, "Pricing and resource allocation in transportation and public utilities," *American Economic Review*, vol. 53, no. 2, pp. 452–465, 1963.
- [4] Y. Sheffi, "Urban transportation networks: equilibrium analysis with mathematical programming methods," *Transportation Research Part A: General*, vol. 20, no. 1, pp. 76–77, 1986.
- [5] P. Ferrari, "Road pricing and network equilibrium," *Transportation Research Part B: Methodological*, vol. 29, no. 5, pp. 357–372, 1995.
- [6] F. Leurent, "Cost versus time equilibrium over a network," *European Journal of Operational Research*, vol. 71, no. 2, pp. 205–221, 1993.
- [7] H. Yan and W. H. K. Lam, "Optimal road tolls under conditions of queueing and congestion," *Transportation Research Part A: Policy and Practice*, vol. 30, no. 5, pp. 319–332, 1996.
- [8] H. Yang and M. G. H. Bell, "Traffic restraint, road pricing and network equilibrium," *Transportation Research Part B: Methodological*, vol. 31, no. 4, pp. 303–314, 1997.
- [9] W. Vickrey, "Congestion theory and transport investment," *American Economic Review*, vol. 59, pp. 251–261, 1969.
- [10] H. Yang and Q. Meng, "Departure time, route choice and congestion toll in a queueing network with elastic demand," *Transportation Research Part B: Methodological*, vol. 32, no. 4, pp. 247–260, 1998.
- [11] M. Carey and D. Watling, "Dynamic traffic assignment approximating the kinematic wave model: System optimum, marginal costs, externalities and tolls," *Transportation Research Part B: Methodological*, vol. 46, no. 5, pp. 634–648, 2012.
- [12] X. Ban and H. X. Liu, "A link-node discrete-time dynamic second best toll pricing model with a relaxation solution algorithm," *Networks and Spatial Economics*, vol. 9, no. 2, pp. 243–267, 2009.
- [13] J. A. Laval and F. Castrillón, "Stochastic approximations for the macroscopic fundamental diagram of urban networks," *Transportation Research Part B: Methodological*, vol. 81, pp. 904–916, 2015.
- [14] N. Geroliminis and C. F. Daganzo, "Macroscopic modeling of traffic in cities," in *Transportation Research Board 86th annual meeting (No. 07-0413)*, Washington D.C., Jan 13–17, 2007.
- [15] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.

- [16] L. Leclercq and N. Geroliminis, "Estimating MFDs in simple networks with route choice," *Procedia-Social and Behavioral Sciences*, vol. 80, pp. 99–118, 2013.
- [17] N. Geroliminis and J. Sun, "Properties of a well-defined macroscopic fundamental diagram for urban traffic," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 605–617, 2011.
- [18] N. Geroliminis, J. Danés, and M. Á. Estrada, "Multimodal Macroscopic Fundamental Diagram for "Car-Bus" Mixed Traffic Signalized Corridors: Application in City of Barcelona," in *The Transportation Research Board 92nd Annual Meeting (No. 13-1089)*, Mixed Traffic Signalized Corridors, Washington D.C., Jan 13-17, 2013.
- [19] Y. Ji and N. Geroliminis, "On the spatial partitioning of urban transportation networks," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1639–1656, 2012.
- [20] D. Sun, K. Zhang, and S. Shen, "Analyzing spatiotemporal traffic line source emissions based on massive didi online car-hailing service data," *Transportation Research Part D: Transport and Environment*, vol. 62, pp. 699–714, 2018.
- [21] K. Zhang, D. J. Sun, S. Shen, and Y. Zhu, "Analyzing spatiotemporal congestion pattern on urban roads based on taxi GPS data," *Journal of Transport and Land Use*, vol. 10, no. 1, pp. 675–694, 2017.
- [22] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, "Exploiting the fundamental diagram of urban networks for feedback-based gating," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1393–1403, 2012.
- [23] K. Aboudolas and N. Geroliminis, "Perimeter and boundary flow control in multi-reservoir heterogeneous networks," *Transportation Research Part B: Methodological*, vol. 55, pp. 265–281, 2013.
- [24] S. Chen and D. J. Sun, "An Improved Adaptive Signal Control Method for Isolated Signalized Intersection Based on Dynamic Programming," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 4, pp. 4–14, 2016.
- [25] Y. Zhang, D. J. Sun, and A. Kondyli, "An empirical framework for intersection optimization based on uniform design," *Journal of Advanced Transportation*, vol. 2017, 2017.
- [26] N. Geroliminis and D. M. Levinson, "Cordon pricing consistent with the physics of overcrowding," in *Transportation and Traffic Theory*, pp. 219–240, Springer, Golden Jubilee, 2009.
- [27] N. Zheng, R. A. Waraich, K. W. Axhausen, and N. Geroliminis, "A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 8, pp. 1291–1303, 2012.
- [28] D. J. Sun and S. Guan, "Measuring vulnerability of urban metro network from line operation perspective," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 348–359, 2016.
- [29] H. Liu, L. Zhang, D. Sun, and D. Wang, "Optimize the Settings of Variable Speed Limit System to Improve the Performance of Freeway Traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3249–3257, 2015.
- [30] D. Sun, Y. Zhang, R. Xue, and Y. Zhang, "Modeling carbon emissions from urban traffic system using mobile monitoring," *Science of the Total Environment*, vol. 599–600, pp. 944–951, 2017.
- [31] H. Yang and H. J. Huang, *Mathematical and Economic Theory of Road Pricing*, Elsevier, Oxford, UK, 2005.
- [32] J. A. Laval, H. W. Cho, J. C. Muñoz, and Y. Yin, "Real-time congestion pricing strategies for toll facilities," *Transportation Research Part B: Methodological*, vol. 71, pp. 19–31, 2015.
- [33] A. H. Meimand, *Differential Stackelberg Games and their Application to Dynamic Pricing, Production Planning Network Design, and Logistics [Ph.D. thesis]*, The Pennsylvania State University, 2013.
- [34] J. M. Thomson, "Toward better urban transport planning in developing countries," *World Bank Staff Working Paper*, vol. 600, 1983.
- [35] X.-Y. Ni, D. Sun, and Z.-R. Peng, "An improved incremental assignment model for parking variable message sign location problem," *Journal of Advanced Transportation*, vol. 49, no. 7, pp. 817–828, 2015.
- [36] V. Gayah and V. Dixit, "Using mobile probe data and the macroscopic fundamental diagram to estimate network densities," *Transportation Research Record*, no. 2390, pp. 76–86, 2013.
- [37] J. Ortigosa, M. Menendez, and H. Tapia, "Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich," *Journal on Transportation Logistics*, vol. 3, no. 3, pp. 245–266, 2014.
- [38] J. Ortigosa, M. Menendez, and V. V. Gayah, "Analysis of network exit functions for various urban grid network configurations," *Transportation Research Record*, vol. 2491, pp. 12–21, 2015.
- [39] L. Leclercq, C. Parzani, V. L. Knoop, J. Amourette, and S. P. Hoogendoorn, "Macroscopic traffic dynamics with heterogeneous route patterns," *Transportation Research Part C: Emerging Technologies*, vol. 59, pp. 292–307, 2015.
- [40] C. F. Daganzo, V. V. Gayah, and E. J. Gonzales, "Macroscopic relations of urban traffic variables: bifurcations, multivaluedness and instability," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 278–288, 2011.
- [41] N. Zheng, G. Rerat, and N. Geroliminis, "A time-dependent area-based pricing scheme for multimodal urban networks with user adaptation: An agent-based approach," in *Proceedings of the 2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, pp. 471–476, China, October 2014.
- [42] K. W. Axhausen, S. Hess, A. Koenig, J. Bates, M. Bierlaire, and G. Abay, "State-of-the-Art Estimates of Swiss Value of Travel Time Savings. the Transportation Research Board 86th Annual Meeting," *Washington D.C.*, 2007.
- [43] Z. Liu, Q. Meng, and S. Wang, "Speed-based toll design for cordon-based congestion pricing scheme," *Transportation Research Part C: Emerging Technologies*, vol. 31, pp. 83–98, 2013.
- [44] Y. Huang, D. Sun, and L. Zhang, "Identifying Intrapersonal Heterogeneity of Driver Behavior from Taxi GPS Traces," *Accident Analysis & Prevention*, vol. 117, pp. 318–327, 2018.
- [45] X. Ni and D. Sun, "Agent-based modelling and simulation to assess the impact of parking reservation system," in *Journal of Advanced Transportation*, vol. 37, pp. 1–10, 2017.
- [46] D. Sun, X. Ni, and L. Zhang, "A discriminated release strategy for parking variable message sign display problem using agent-based simulation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 38–47, 2015.
- [47] P. Johnson, D. Nguyen, and M. Ng, "Large-scale network partitioning for decentralized traffic management and other transportation applications," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 20, no. 5, pp. 461–473, 2016.

Research Article

Automatic Estimation Method for Intersection Saturation Flow Rate Based on Video Detector Data

Linhong Wang ¹, Yunhao Wang,² and Yiming Bie ¹

¹School of Transportation, Jilin University, Changchun 130022, China

²School of Economics, Northeast Normal University, Changchun 130117, China

Correspondence should be addressed to Yiming Bie; yimingbie@126.com

Received 30 April 2018; Revised 20 June 2018; Accepted 5 July 2018; Published 17 July 2018

Academic Editor: David Z. W. Wang

Copyright © 2018 Linhong 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.

Saturation flow rate (SFR) is a fundamental parameter to the level of service evaluation, lane capacity calculation, and signal timing plan optimization at signalized intersections. It is affected by a variety of factors including weather conditions, lane width, and the type of the driver. How to accurately estimate the SFR remains one of the most important tasks in traffic engineering. Existing studies generally rely on the field measurement method which requires a large number of people collecting data at the intersection. As a result, the method incurs a high economic cost and cannot adapt to the dynamic change of SFR. In recent years, video detectors have been widely installed at intersections which are capable of recording the time each vehicle passes the stop line, the number plate of each vehicle, and the vehicle type. This paper therefore aims to propose an automatic estimation method for the SFR based on video detector data in order to overcome the limitation of the field measurement method. A prerequisite for estimating the SFR is to recognize the saturation headway. We consider the actual vehicle headway as time series and build an auxiliary regression equation whose parameters are estimated through the ordinary least squares method. We employ the Dickey-Fuller test to verify whether the headways in the time series are saturation headways. An iterative method using quantiles is proposed to filter out abnormal data. The SFR is finally calculated using the average value of saturation headways. To demonstrate the proposed method, we conduct a case study using data from an intersection with three entrance lanes in Qujing city, Yunnan Province, China. The overall estimation process is displayed and the impacts of quantile selection and data duration on the estimation accuracy are analyzed.

1. Introduction

In traffic engineering domain, the saturation flow rate (SFR) of an approaching lane at the intersection is a fundamental parameter for signal timing optimization, capacity calculation, level of service evaluation, and intersection design. As defined by the Highway Capacity Manual (HCM), the SFR is the flow in vehicles per hour that can be accommodated by the lane group assuming that the green phase was displayed 100 percent of the time [1]. The SFR cannot be measured directly and hence needs to be estimated by traffic engineers. While different intersections may have different SFRs, the same intersection may also exhibit a changing SFR because it is affected by a variety of factors including weather conditions, lane width and slope, the type of the driver, and the ratio of heavy vehicles [2, 3].

Given the importance of the SFR, a number of studies have been devoted to its estimation methods. Two methods are provided by HCM. The first method is called the adjustment method. Computations begin with the selection of a base SFR, usually 1900 passenger cars per hour per lane (pcu/h/ln). This value is adjusted for a variety of factors, such as number of lanes in lane group, lane width, heavy vehicles, approach grade, parking, bus blockage, type of area, lane utilization, left turns, right turns, pedestrian, and pedestrian-bicycle. The adjustment formula for each factor is given. However, adjusting the SFR considering such a variety of factors would result in a huge amount of work [4]. HCM therefore provides the second method which is called the field measurement method. Field-measured values of SFR will produce more accurate results than the estimation procedure described in the adjustment method and can be used directly

without further adjustment. SFR is the maximum discharge rate during the green time. It is usually achieved after about 10 to 14 s of green, which corresponds to the front axle of the fourth to sixth passenger car crossing the stop line after the beginning of green. The period of saturation flow begins when the front axle of the fourth vehicle in the queue crosses the stop line or reference point and ends when the front axle of the last queued vehicle crosses the stop line. The last queued vehicle may be a vehicle that joined the queue during the green time. The times when the fourth to the last queued vehicle cross the stop line are recorded and used to calculate the average headway per vehicle. The SFR is 3,600 divided by this value. In order to obtain a statistically significant value, a minimum of 15 signal cycles with more than eight vehicles in the initial queue is typically required. An average of the SFR values in individual cycles represents the prevailing local SFR for the surveyed lane [5, 6].

Shanteau [7] proposed a method to estimate SFR based on cumulative curve method. In the cumulative curve figure, the horizontal axis represented time and the vertical axis represented the number of the vehicle crossing the stop line. A diagram was presented to show the average number of vehicles to enter the intersection versus time. Then a straight line was selected to best fit the data. The straight line portion of the plot could be determined by eye (the easiest and usually sufficient way) or by some formalized method such as linear regression. Then the SFR could be determined according to the slope of the straight line. Hossain [8] developed a microsimulation model to depict the traffic flows in developing cities, where the traffic comprised both motorized and nonmotorized vehicles. The saturation flows at signalized intersections were investigated in an aggregate form of vehicles per hour. It was found that the saturation flows were dependent on the approach width, turning proportion, and composition of the traffic mix. A regression model was proposed to estimate the saturation flow from the influencing variables. Shao and Liu [9] found that using the average value of observed queue discharge headways to estimate the saturation headway might lead to underestimate SFR. Based on the field data, normal and log-normal distributions were used to fit the headways. The median value of discharge headways was suggested to estimate the saturation headway. Zhao et al. [10] studied the headway distribution function at signalized intersections using field collected data. The relationship between headway and vehicle position were modeled and the 78th percentile of headway at each position was selected to calibrate the SFR.

The SFR is equal to 3,600 divided by the saturation headway. When the saturation headway is obtained, the SFR is automatically determined. Therefore, a few methods for estimating the saturation headway have been proposed. Tong and Hung [11] proposed a neural network approach to simulate the queued vehicle discharge headway. A computer-based three-layered neural network model was developed for the estimation of discharge headway. Hung et al. [12] found that discharge headway of individual vehicles at different queue position follows the Type I Extreme Value Distribution. They proposed a method to estimate site-specific parameters for this distribution. Murat and Gedizlioglu [13] investigated the

vehicle time headways at signalized intersection using the data in Turkey. Seven distribution models were taken into account to fit the headway data and log-normal, Weibull, gamma, or Pearson distribution models performed better than other models. Jin et al. [14] found that distributions of the departure headways at each position (except the first position) in a queue were revealed to approximately follow a certain log-normal distribution. Then they proposed a new car-following model from the driver's perspective to explain why the departure headways follow such a log-normal distribution. Tan et al. [15] compared the departure headway collected from intersections in Beijing and generated by VISSIM and TransModeler. Results showed that slight difference in car-following models and parameters would result in significant difference of the distribution. Yin et al. [16] compared the departure headways in Beijing and Atlanta. Urban traffic flows in Beijing, China, were often hindered by pedestrians at intersections and queuing vehicles often cannot be totally discharged during the green time. It was found that the departure headways of the first and second vehicles and after the 5th vehicle in Beijing were significantly different from those in Atlanta. Tan et al. [17] developed two distribution models to fit startup lost time and effective flow rate, respectively. The dependence of departure headways in three types of simulation software (VISSIM, PARAMICS, and TransModeler) was also examined. Hao and Ma [18] tried to validate the assumption that the mean headway of saturation flow was saturated headway. They investigated the distribution of departure headways based on the field data in Shanghai, China. It was found that shifted log-normal was more suitable to capture the distribution of steady-flow headways, and the mode of steady-flow headway distribution could be considered as theoretical saturation headway and the mean could be considered as practical saturation headway.

While quite some research effort has been made to estimate the intersection SFR, all the aforementioned methods are essentially the field measurement method provided by HCM. These methods require traffic engineers to conduct filed surveys at intersections. The number of vehicles queuing behind the stop line is recorded manually and the times when the fourth to the last queued vehicle cross the stop line are extracted to calculate the saturation headways. Although the field measurement method can estimate the saturation headway and hence the SFR, it has the following two limitations:

(i) High labor and economic costs: as suggested by HCM, two surveyors are needed to estimate the SFR for one entrance lane [1]. While the field measurement method may be feasible when the number of entrance lanes is small, it can hardly be applied in a city where there are hundreds of intersections and hence thousands of entrance lanes. Therefore, to be able to estimate the SFR, a large number of surveyors are required to collect data for each lane resulting in a high economic cost.

(ii) Inability to adapt to the dynamic change in SFR: traffic data collected by the field measurement method and the resulting estimations of the SFR only represent a single or a few typical traffic scenarios. Since the SFR is affected by a number of factors, there can be hundreds of traffic scenarios and, hence, the method can hardly be applied to estimate

the SFR under all possible scenarios. Moreover, the real-time SFR is an input parameter for adaptive signal control at intersections. When one of the influencing factors changes (e.g., weather or the ratio of heavy vehicles), the SFR changes accordingly. While it may take at least 30 minutes to send traffic engineers to the site and obtain the adjusted SFR, this obviously cannot fulfill the real-time requirement of adaptive signal control.

In recent years, video detectors have been widely installed at signalized intersections in many cities. They can provide fruitful traffic information to advanced traffic management system, such as traffic volumes of different types of vehicles, volume of pedestrian, vehicle trajectory, and traffic OD (origin-destination) matrix. For example, with the help of video detectors, adaptive traffic control system can be developed to adjust signal timing plans in real time and estimate the queue length behind the stop line. In addition, video detectors can also provide time headways by recording in real time the number plate and time of each vehicle crossing the stop line and the type of the vehicle. The advantages of such an automatic estimation method for the SFR using video detector data include the: (i) vehicle discharge headways during the green phase can be automatically recorded without the need for manual surveys, which reduces the economic cost; (ii) vehicle operating data under different traffic scenarios can be recorded in real time, which fulfills the real-time requirement of traffic control; (iii) the method can be applied to the entire urban road network to automatically estimate the SFR for each entrance lane under each traffic scenario.

However, vehicle discharge headways obtained from video detector data cannot be used directly to estimate the SFR. This is because, as shown in HCM, the SFR can only be estimated using the saturation headways from the queuing vehicles. While the field measurement method allows traffic engineers to identify the queuing vehicles on-site, video detectors only record the time each vehicle crosses the stop line rather than automatically identifying the queuing vehicles. Therefore, further analysis on video detector data is needed.

Different video detectors can record different traffic flow parameters. The video detectors used in this study cover the area surrounding the stop line and record the following parameters: the number plate, speed, direction, and time of each vehicle crossing the stop line, the type of the vehicle, and the number of the entrance lane. Such a technique has already been widely applied at intersections in a number of cities in China. Therefore, the objective of this study is to propose an automatic estimation method for the intersection SFR based on video detector data that can overcome the limitations of the field measurement method (i.e., high labor cost and inability to fulfill the real-time requirement of traffic control). The rest of the paper is organized as follows. Section "Automatic Estimation Method for Saturation Flow Rate" proposes the automatic estimation methodology, the core of which is the automatic recognition of the saturation flow. Section "Case Study" conducts a case study on a field intersection and analyzes the effects of different parameters on the SFR. The last section concludes the paper.

2. Automatic Estimation Method for Saturation Flow Rate

2.1. Recognition of Saturation Flow. As described in the HCM, the period of saturation flow begins when the front axle of the fourth vehicle in the queue crosses the stop line or reference point and ends when the front axle of the last queued vehicle crosses the stop line. The first step of estimating the SFR is to recognize the saturation flow during the green phase, i.e., to recognize the fourth to the last queued vehicles. Since video detectors record the time each vehicle crosses the stop line, it is rather easy to obtain the time the fourth vehicle crosses the stop line. However, it is impossible to tell which vehicle is the last in the queue because video detectors do not record the real-time queue length. Therefore, a new method is needed to recognize the saturation flow.

During the green phase, the queuing vehicles cross the stop line successively. If the traffic flow is saturated, vehicle headways should fluctuate around the saturation headway. Therefore, from a statistics perspective, vehicle headways should fluctuate around a certain constant and the fluctuations are irregular. In this sense, the error series of the vehicle headway series should be independently and identically distributed (IID) with zero mean. Let X_m denote the time series of all sampled saturation headways during the peak period, with observations denoted by $X_m(1), X_m(2), \dots, X_m(n), \dots, X_m(N_m)$, where $1 \leq n \leq N_m$ and N_m is the number of samples. The n -th time headway $X_m(n)$ is therefore expressed as follows:

$$X_m(n) = \mu + e(n) \quad (1a)$$

where μ is the true saturation headway and $e(n)$ is the random error term of the n -th headway. We assume that e follows a normal distribution with zero mean and variance σ_e^2 and that any two of the random errors are IID. Since μ is unknown and to be estimated, the fluctuations of the observed headways around μ are unknown. Thus, σ_e^2 is also unknown.

Since e follows a normal distribution and μ is a constant, the expectation and variance of X_m can be obtained based on the following two equations:

$$E[X_m(n)] = \mu \quad (1b)$$

$$\text{var}[X_m(n)] = \sigma_e^2 \quad (1c)$$

That is, the saturation headway series X_m follows a normal distribution with mean equal to μ and variance equal to σ_e^2 .

According to (1a), the $(n-1)$ -th time headway in X_m can be expressed as follows:

$$X_m(n-1) = \mu + e(n-1) \quad (1d)$$

Both (1a) and (1d) contain the real saturation flow rate μ ; thus we can subtract (1d) from (1a) to eliminate μ and then obtain the following relation:

$$X_m(n) - X_m(n-1) = e(n) - e(n-1) \quad (2a)$$

TABLE 1: Critical values of the *DF* distribution.

Number of samples	Quantile α							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
25	-2.66	-2.26	-1.95	-1.60	0.92	1.33	1.70	2.16
50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.08
100	-2.60	-2.24	-1.95	-1.62	0.90	1.29	1.64	2.03
250	-2.58	-2.23	-1.95	-1.62	0.89	1.29	1.63	2.01
500	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00
∞	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00

By setting $\Delta e(n) = e(n) - e(n-1)$ and $\Delta X_m(n) = X_m(n) - X_m(n-1)$, (2a) is further simplified:

$$\Delta X_m(n) = \Delta e(n) \quad (2b)$$

Let Δe and ΔX_m denote the time series comprised $\Delta e(n)$ and ΔX_m , respectively. Since any two of the random errors are IID, the expectation and variance of $\Delta e(n)$ are expressed as follows:

$$\begin{aligned} E[\Delta e(n)] &= E[e(n) - e(n-1)] \\ &= E[e(n)] - E[e(n-1)] = 0 \end{aligned} \quad (3a)$$

$$\begin{aligned} \text{var}[\Delta e(n)] &= \text{var}[e(n) - e(n-1)] \\ &= \text{var}[e(n)] + \text{var}[e(n-1)] = 2\sigma_e^2 \end{aligned} \quad (3b)$$

Equations (3a) and (3b) show that Δe follows a normal distribution with zero mean and $2\sigma_e^2$ variance, which provides a first method to test whether the vehicle headways in the time series are saturation headways. Specifically, we first obtain ΔX_m and Δe through the observed X_m and calculate the sample mean and variance. We then perform a hypothesis test of the population mean (i.e., *t*-test with unknown population variance) by checking whether there is a significant difference between the sample mean and zero. If there is no significant difference, the expectation of Δe is considered zero. While the sample variance of Δe can be calculated, σ_e^2 cannot be obtained through (1a). As a result, we cannot check whether the sample variance of Δe is equal to $2\sigma_e^2$ and the method cannot work properly.

To better test whether the observed time headways are saturation headways, we set $k(n) = \Delta e(n)$ and $\sigma_k^2 = 2\sigma_e^2$ and propose the following auxiliary regression equation:

$$X_m(n) = \rho X_m(n-1) + k(n) \quad (4a)$$

where ρ is a parameter to be estimated.

After (4a) has been developed, the next step is to perform a hypothesis test with the null hypothesis of $\rho = 1$. If the null hypothesis cannot be rejected, the observed headways can be considered as saturation headways. In such condition, $X_m(n) = X_m(n-1) + k(n)$. Because $k(n) = \Delta e(n)$ and $E[\Delta e(n)] = E[k(n)] = 0$ (please refer to (3a)), then we can get $E[X_m(n)] = E[X_m(n-1)]$. Thus, it can be concluded that the headways in X_m are saturation headways.

The parameters ρ and $k(n)$ in (4a) are estimated using the ordinary least squares (OLS) method. The resulting estimations are denoted by $\hat{\rho}$ and $\hat{k}(n)$ as follows:

$$\hat{\rho} = \frac{\sum_{n=2}^{N_m} X_m(n-1) X_m(n)}{\sum_{n=2}^{N_m} [X_m(n-1)]^2} \quad (4b)$$

$$\hat{k}(n) = X_m(n) - \hat{\rho} X_m(n-1) \quad (4c)$$

According to classical regression theory, $(\hat{\rho} - \rho)/SE(\hat{\rho})$ should follow a *t* distribution with a $(N_m - 2)$ degree of freedom. If $\rho = 1$ is satisfied, $(\hat{\rho} - 1)/SE(\hat{\rho})$ should follow a *t* distribution with a $(N_m - 2)$ degree of freedom where $SE(\hat{\rho})$ is the standard error of $\hat{\rho}$. A smaller standard error indicates more closeness between the sample statistics and the population parameters. The samples are therefore a better representative of the population which results in more reliable estimations of the population parameters.

$$SE(\hat{\rho}) = \frac{\hat{\sigma}_k^2}{\sum_{n=2}^{N_m} [X_m(n-1) - \bar{X}_m]^2} \quad (4d)$$

$$\bar{X}_m = \frac{1}{N_m - 1} \sum_{n=2}^{N_m} X_m(n-1) \quad (4e)$$

$$\hat{\sigma}_k^2 = \frac{1}{N_m - 2} \sum_{n=2}^{N_m} [\hat{k}(n)]^2 \quad (4f)$$

Since (4a) involves both the *n*-th and (*n*-1)-th vehicle headways, the characteristics of the constructed statistic for the hypothesis test of ρ may change. It is proved by Dickey and Fuller [19] that when a hypothesis test is conducted on ρ in (4a), the constructed “*t*” statistic based on the null hypothesis of $\rho = 1$ no longer follows the *t* distribution. Instead, it follows the *DF* distribution. To distinguish between the two statistics, we construct the following *DF* statistic:

$$DF = \frac{(\hat{\rho} - 1)}{SE(\hat{\rho})} \quad (4g)$$

The constructed *DF* statistic in (4g) follows the *DF* distribution whose critical values are needed during the hypothesis test. Table 1 shows the critical values of the *DF* distribution corresponding to different quantiles.

Table 1 shows that when we perform a two-sided test with a significance level of 5%, the acceptance region lies

between -2.25 and 1.66 given that the number of samples is larger than 50. That is, when the DF statistic lies within $[-2.25, 1.66]$, the null hypothesis of $\rho = 1$ cannot be rejected. The observed headways can be considered as saturation headways. In statistics, the time series X_m satisfying $X_m(n) = X_m(n-1) + k(n)$ is termed a random walk.

2.2. Exclusion of Abnormal Data. While section “Recognition of Saturation Flow” proposes a method to test whether the observed headways are saturation headways, this section aims to propose a method to filter out abnormal data.

Estimating the SFR usually needs traffic data collected during a peak period. Let t_{start} and t_{end} denote the beginning and ending times of the peak period, respectively. Based on the outputs of video detectors (i.e., the time each vehicle crosses the stop line), we can obtain the time series of the observed headways within $[t_{start}, t_{end}]$, denoted by X_0 . The number of samples is N_0 and the n -th sample is $X_0(n)$ where $1 \leq n \leq N_0$. When we construct X_0 , various factors that may affect the headway are not taken into account. As a result, the headways in X_0 may not necessarily be at saturation.

There are three main types of abnormal headways in X_0 : (a) due to the red light, the headway may be large between the first vehicle crossing the stop line in current cycle and the last vehicle crossing the stop line in the previous cycle; (b) at the beginning of the green phase, the headways between the first few vehicles crossing the stop line may be large because they are still accelerating; (c) during the green phase, there may be a large headway as a result of the inattention of drivers or the random arrivals of vehicles.

When the calculated headway is larger than or equal to the red time of the signal phase, the (a) type of headways occurs and is excluded from X_0 . The time series after the exclusion is denoted by X_1 and the number of samples is N_1 . Since X_1 still includes the (b) and (c) types of headways, the next step is to find a threshold θ whereby the headways smaller than θ are considered as saturation headways (i.e., through the DF test shown in section “Recognition of Saturation Flow”).

Since X_1 is calculated using traffic data collected during a peak period, the number of abnormal samples of types (b) and (c) only accounts for a small proportion of all the samples, and the majority of observed headways are fluctuating around the saturation headway. The mode of the sampled data in X_1 remains around the saturation headway rather than becoming larger due to abnormal data. If the mode exists for X_1 containing abnormal data, the mean value should be larger than the mode; i.e., X_1 is a right-skewed distribution. As a result, using the mean value of X_1 as the threshold θ is inappropriate. Using the mode as the threshold is, however, also inappropriate because, under some circumstances, the mode does not exist for X_1 . Moreover, the mode of some headway samples may be smaller than the saturation headway. Therefore, this paper uses sample quantiles to set the threshold θ .

When the number of samples is large, we can use the 60%-80% quantile to set θ which excludes more data from X_1 . When the number of samples is small, we can use the 80%-95% quantile as the threshold which excludes fewer data from

X_1 . The detailed steps of filtering out abnormal data are as follows.

Step 1. Initialize and set $i = 1$.

Step 2. Calculate the headway corresponding to the quantile β in X_i and set as the threshold θ .

Step 3. Exclude data larger than θ from X_i and generate the new time series X_{i+1} (the number of samples becomes N_{i+1}). Conduct a DF test on X_{i+1} . If the test passes, terminate the iterative algorithm; otherwise go to Step 4.

Step 4. Set $i = i + 1$ and go to Step 2.

The time series generated at the last iteration is denoted by X_m consisting of N_m samples. The algorithmic steps show that each iteration reduces the number of samples by a factor of β ($\beta < 1$) and that the headway corresponding to the quantile β changes accordingly. The number of samples in X_m , i.e., N_m , is equal to $\beta^{m-1} \times N_1$.

2.3. Calculation of the Saturation Flow Rate. From section “Exclusion of Abnormal Data”, we obtain the time series X_m where the headways are denoted by $X_m(1), X_m(2), \dots, X_m(N_m)$, respectively. Since X_m passes the DF test, it contains saturation headways and should follow a normal distribution as proved by (1a), (1b), (1c), and (1d). In this sense, the mean value, median, and mode of X_m are nearly equal.

We use the sample mean value to estimate the true saturation headway μ . The estimation $\hat{\mu}$ is expressed as follows:

$$\hat{\mu} = \frac{1}{N_m} \sum_{n=1}^{N_m} X_m(n) \quad (5a)$$

By taking the expectation on both sides of (5a) and plugging it into (1b), we obtain $E[\hat{\mu}] = \mu$. Therefore, $\hat{\mu}$ is an unbiased estimation of μ and the estimation accuracy can be expressed by $\text{var}[\hat{\mu}]$ where $\text{var}[\hat{\mu}] = \sigma_{\hat{\mu}}^2/N_m$. $\sigma_{\hat{\mu}}^2$ is the unknown variance of the true saturation headway μ . We can use the variance of X_m , i.e., $S_{\hat{\mu}}^2$, as the estimation of $\sigma_{\hat{\mu}}^2$:

$$S_{\hat{\mu}}^2 = \frac{1}{N_m - 1} \sum_{n=1}^{N_m} [X_m(n) - \hat{\mu}]^2 \quad (5b)$$

The interval estimation of μ with a significance level of α is as follows:

$$\begin{aligned} \hat{\mu} - t_{\alpha/2}(N_m - 1) \frac{S_{\hat{\mu}}^2}{\sqrt{N_m}} &\leq \mu \\ &\leq \hat{\mu} + t_{\alpha/2}(N_m - 1) \frac{S_{\hat{\mu}}^2}{\sqrt{N_m}} \end{aligned} \quad (6)$$

where $t_{\alpha/2}(N_m - 1)$ is the critical value of the t distribution with a $(N_m - 1)$ degree of freedom and a significance level of α .

The limit error of the estimated μ , i.e., δ , is expressed as follows:

$$\delta = t_{\alpha/2}(N_m - 1) \frac{S_{\hat{\mu}}^2}{\sqrt{N_m}} \quad (7a)$$

TABLE 2: Statistics of the raw time headways.

Statistical parameters	Lane 1	Lane 2	Lane 3
Number of samples	810	1247	986
Mean	11.029	8.634s	9.526s
Median	3.063s	2.521s	2.874
Mode	2.867s	2.400s	2.704s
Standard deviation	15.292	12.619s	18.835s
Minimum	1.597s	1.259s	1.332s
Maximum	163.006s	145.572s	145.384s

Assuming that the minimum number of samples to ensure that δ is smaller than ε is N_ω , we obtain the following relation:

$$t_{\alpha/2}(N_\omega - 1) \frac{S_\mu^2}{\sqrt{N_\omega}} \leq \varepsilon \quad (7b)$$

We can use (7b) to solve the interval for N_ω . The minimum value of N_ω is achieved when the error precision ε is reached. When the number of samples N_m is much larger than N_ω , the estimated SFR using the mean value of the samples can achieve a very high precision. If N_m is smaller than N_ω , we can use headway data in peak hours of multiple days (e.g., a week or a month) to increase N_m .

3. Case Study

We apply the proposed method to a signalized intersection in Qujing, China, and present the overall estimation process of the SFR. Qujing is the second largest city in Yunan province, with the main urban area of about 130 square kilometers and a population of 1.0 million. The intersection is made up of Qilin South Road and Wenchang Street and the layout is shown in Figure 1. There are three approaching lanes and three exiting lanes on each leg. To illustrate the estimation of SFR, the three approaching lanes on the north leg are selected as examples: lane 1 dedicated to left turn vehicles, lane 2 for through vehicles, and lane 3 shared by through and right-turn vehicles. The lane width is 3.5 meters. Video detectors are installed for each lane recording data such as the number plate, type, and time of each vehicle crossing the stop line. The cycle length of the signal timing during the morning peak is 180 seconds. The green time, red time, and amber time for lane 1 are 19 seconds, 158 seconds, and 3 seconds, respectively. The green time, red time, and amber time for lanes 2 and 3 are the same because they are controlled by the same signal phase, which are 36 seconds, 141 seconds, and 3 seconds, respectively.

3.1. Data Description. We use data collected during the morning peak (7:00 – 10:00) on October 30, 2017, to estimate the SFRs for the three lanes. The weather was fine without rains. The numbers of raw samples of headways for the three lanes are 810, 1,247, and 986, respectively. Table 2 shows the resulting statistics. It is obvious that the raw samples contain quite a few abnormal data. For example, the maximum headway of lane 2 reaches 145.572 seconds, which actually

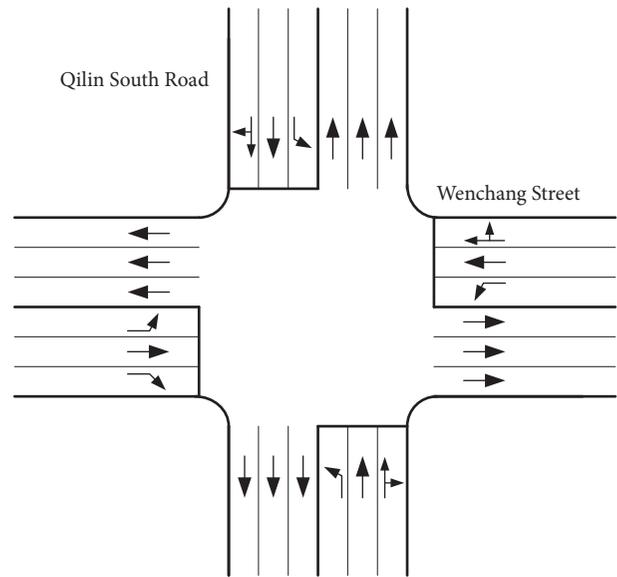


FIGURE 1: The layout of the survey intersection.

represents the headway between the first vehicle crossing the stop line in the current cycle and the last vehicle crossing the stop line in the previous cycle.

3.2. Saturation Flow Estimation Using the 80% Quantile. We first use the red time for each lane as the threshold and exclude the (a) type of abnormal data shown in the section “Exclusion of Abnormal Data”. That is, we exclude headways larger than 158 seconds and 141 seconds for lane 1 and lanes 2 and 3, respectively. The numbers of samples for the three lanes after the exclusion become 743, 1,183, and 922, respectively. Table 3 shows the resulting statistics. By comparing Table 2 with Table 3, we observe that the mean values of the samples reduce after the (a) type of abnormal data is excluded. However, the differences between the mean values, medians, and modes are still large. The maximum values of the samples for the three lanes reach 10.256 s, 8.884 s, and 9.857 s, respectively.

To exclude the (b) and (c) types of abnormal data, we apply the algorithmic steps proposed in Section “Exclusion of Abnormal Data” with $\beta = 80\%$. We take lane 2 as an example and show the process in Table 4 where H_0 represents the null hypothesis. The samples pass the DF test after six

TABLE 3: Statistics of the samples after excluding the (a) type of abnormal data.

Statistical parameters	Lane 1	Lane 2	Lane 3
Number of samples	743	1184	922
Mean	4.643s	3.648s	4.257s
Median	2.937s	2.441s	2.661s
Mode	2.867s	2.400s	2.704s
Standard deviation	7.421s	5.037s	6.633s
Maximum	1.597s	1.259s	1.332s
Minimum	10.256s	8.884s	9.857s

TABLE 4: Applying the iterative method using samples from lane 2.

Number of iteration	Number of samples	DF statistic	Reject or not $H_0 : \rho = 1$	80% quantile (θ)
1	1184	-20.77	Reject	3.73s
2	947	-5.23	Reject	2.88s
3	759	-3.63	Reject	2.49s
4	607	-2.82	Reject	2.28s
5	486	-2.27	Reject	2.16s
6	392	-1.83	Do not reject	/

TABLE 5: Saturation headway estimations based on the 80% quantile for the three lanes.

Data	Lane 1	Lane 2	Lane 3
T.N.I.	3	6	5
N.S.	380	392	302
Final DF statistic	-1.91	-1.83	-1.87
Mean value of SH(s)	2.421	1.790	2.232
Median of SH (s)	2.404	1.801	2.229
Mode of SH (s)	2.419	1.796	2.224
S.D. of SH (s)	0.544	0.251	0.362
Limit error of SH (s)	0.055	0.025	0.041
I.E. of SH (s)	[2.366, 2.476]	[1.765, 1.815]	[2.191, 2.273]
Estimated SFR (pcu/h)	1487	2011	1613
I.E. of SFR (pcu/h)	[1454, 1522]	[1983, 2039]	[1583, 1643]

* T.N.I. represents total number of iterations; N.S. denotes number of the samples; S.D. represents standard deviation; I.E. denotes interval estimation; SH is the abbreviation of saturation headways.

iterations and the number of samples reduces to 392. The mean value, median, and mode of these samples are 1.790 s, 1.801 s, and 1.796 s, respectively, which are nearly equal. This implies that the 392 samples obtained after six iterations follow a normal distribution and that the value of each sample fluctuates around a certain constant (i.e., the true saturation headway).

We therefore use the mean value of the samples to estimate the saturation headway for lane 2; i.e., $\hat{\mu}_2 = 1.790$, where the subscript 2 represents the number of the lane. Hence, the SFR of lane 2 is estimated to be $\hat{\lambda}_2 = 2011$ pcu/h. The standard deviation of the samples is 0.251 s. Given $\alpha = 0.05$ and $t_{\alpha/2}(391) = 1.97$, the limit error of the saturation headway of lane 2 is $\delta_2 = 1.97 \cdot 0.251/\sqrt{392} = 0.025$ s. We are therefore 95% confident that the saturation headway of lane 2 lies within $[\hat{\mu}_2 - \delta_2, \hat{\mu}_2 + \delta_2]$, i.e., [1.765 s,

1.815 s]. The corresponding interval estimation of the SFR is [1983 pcu/h, 2039 pcu/h]. The method is also applied to lanes 1 and 3 and the results are shown in Table 5. The results of lane 2 are included for the comparison.

3.3. Effect of Quantile Selection on the Estimated Saturation Flow Rate. When filtering out abnormal data in section “Saturation Flow Estimation using the 80% Quantile”, we assume $\beta = 80\%$ indicating that the number of samples reduces by 20% during each iteration. In this section, we aim to analyze the effect of different quantiles on the iterative process and the estimated saturation headway.

We choose β to be 60%, 70%, and 90%, respectively, and estimate the corresponding SFRs based on the collected headway samples of lane 2. The results are compared with what we have obtained in section “Saturation Flow Estimation

TABLE 6: SFR estimations based on the 60%, 70%, 80%, and 90% quantiles for lane 2.

Data	$\beta = 60\%$	$\beta = 70\%$	$\beta = 80\%$	$\beta = 90\%$
T.N.I.	3	4	6	12
N.S.	258	407	392	383
Final <i>DF</i> statistic	-1.34	-1.89	-1.83	-1.79
Mean value of SH(s)	1.661s	1.804s	1.790s	1.782s
Median of SH (s)	1.720s	1.810s	1.801s	1.800s
Mode of SH (s)	1.920s	1.805s	1.805s	1.805s
S.D. of SH (s)	0.210s	0.256s	0.251s	0.247s
Limit error of SH (s)	0.026s	0.025s	0.025s	0.025s
I.E. of SH (s)	[1.635, 1.687]	[1.779, 1.829]	[1.765, 1.815]	[1.757, 1.807]
Estimated SFR (pcu/h)	2167	1995	2011	2020
I.E. of SFR (pcu/h)	[2133, 2201]	[1968, 2023]	[1983, 2039]	[1992, 2049]

TABLE 7: SFR estimations of lane 2 under different data durations.

Data duration (min)	30	60	90	120	150	180
N.R.S.	166	399	644	855	1054	1247
T.N.I.	3	4	4	5	6	6
N.S.	99	197	315	335	338	392
Final <i>DF</i> statistic	-1.47	-1.52	-1.92	-1.82	-1.75	-1.83
Mean value of SH(s)	2.276s	1.977s	1.924s	1.838s	1.788s	1.790s
Median of SH (s)	2.285s	2.040s	1.921s	1.880s	1.800s	1.801s
Mode of SH (s)	2.400s	2.400s	1.796s	1.796s	1.796s	1.796s
S.D. of SH (s)	0.459s	0.354s	0.324s	0.280	0.246	0.251
Limit error of SH (s)	0.091	0.049	0.036	0.030	0.027	0.025
I.E. of SH (s)	[2.185, 2.367]	[1.928, 2.026]	[1.888, 1.960]	[1.808, 1.868]	[1.761, 1.815]	[1.765, 1.815]

*N.R.S. denotes number of raw samples.

using the 80% Quantile”, which are further summarized in Table 6.

Table 6 shows that when β is equal to 70%, 80%, or 90%, the mean value, median, and mode of the saturation headways are nearly equal. However, there is a big difference when β is equal to 60% which implies that the saturation headways do not follow a normal distribution. This is because when a small quantile is selected, a large number of samples are excluded by the iterative method which, however, may not all be abnormal data. As a result, the remaining headways tend to be smaller. When excluding the (a) type of abnormal data for lane 2 in Table 3, we are left with 1,184 samples. In Table 6, the numbers of iterations corresponding to β equal to 60%, 70%, 80%, and 90% are 3, 4, 6, and 12, respectively, and the numbers of the remaining samples are 258, 407, 392, and 383, respectively. It is obvious that when β is equal to 60%, a large number of samples are excluded and, hence, a small β is not recommended.

A larger β , however, may result in a larger number of iterations which reduces the real-time capability of the estimation method without increasing its accuracy. Therefore, we suggest β be set at 80%.

3.4. Effect of Data Duration on the Estimated Saturation Flow Rate. While video detectors are capable of continuously recording traffic data for estimating the SFR, excessive data

do not necessarily result in higher accuracy. However, the accuracy may reduce if the number of samples is insufficient. This section therefore aims to analyze the effect of data duration on the estimated SFR.

The data used in the two previous subsections were collected during the 7:00-10:00 morning peak on October 30, 2017. We take lane 2 as an example and estimate the SFRs corresponding to 30 min, 60 min, 90 min, 120 min, 150 min, and 180 min of data duration. The estimated SFRs are summarized in Table 7 with β equal to 80%.

Table 7 shows that when the data duration is small, the mean value, median and mode of the saturation headways exhibit a large difference which violates the assumption of a normal distribution. When the data duration increases to 150 min and 180 min, all the three values are nearly equal. Therefore, we suggest using the data duration no less than 150 min to ensure the estimation accuracy and robustness.

4. Conclusion

By using video detector data, this study proposes an automatic estimation method for the intersection SFR consisting of the following three sections: automatic recognition of the saturation flow, exclusion of abnormal data, and calculation of the SFR based on the saturation headway. The following conclusions are made.

(i) Recognition of saturation flow is the key to the estimation of the SFR. In this study, saturation flow is recognized by constructing the DF statistic and conducting the DF test. We find that the time series of the saturation headways is a random walk in statistics.

(ii) While headways can be obtained through video detector data, there exist quite a few abnormal data that need to be filtered out to ensure the estimation accuracy. The iterative method based on quantiles proposed in this study is able to exclude these abnormal data and maintain a high level of flexibility.

(iii) Both quantile selection and data duration have an impact on the estimation accuracy of the SFR. We suggest using the quantile of 80% and the data duration of no less than 150 min.

Data Availability

The data supporting the conclusions of the study can be found in the section of "Case Study".

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (nos. 71771062, 71771050, and 71501038), the Postdoctoral Startup Foundation of Heilongjiang Province (LBH-Q15050), and the Opening Foundation of the Intelligent Traffic Engineering Technology Research Center of Yunan Province Universities.

References

- [1] Transportation Research Board, "Highway Capacity Manual," *National Research Council*, 2000.
- [2] S. Wang and X. Qu, "Rural bus route design problem: Model development and case studies," *KSCE Journal of Civil Engineering*, vol. 19, no. 6, pp. 1892–1896, 2015.
- [3] Y. Bie, X. Gong, and Z. Liu, "Time of day intervals partition for bus schedule using GPS data," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 443–456, 2015.
- [4] Z. Liu, S. Wang, W. Chen, and Y. Zheng, "Willingness to board: a novel concept for modeling queuing up passengers," *Transportation Research Part B: Methodological*, vol. 90, pp. 70–82, 2016.
- [5] Z. Liu, S. Wang, B. Zhou, and Q. Cheng, "Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 58–72, 2017.
- [6] Z. Liu, Y. Yan, X. Qu, and Y. Zhang, "Bus stop-skipping scheme with random travel time," *Transportation Research Part C: Emerging Technologies*, vol. 35, pp. 46–56, 2013.
- [7] J. Bared, G. L. Giering, and D. L. Warren, "Safety evaluation of acceleration and deceleration lane lengths," *ITE Journal*, vol. 69, no. 5, pp. 50–54, 1999.
- [8] M. Hossain, "Estimation of saturation flow at signalised intersections of developing cities: a micro-simulation modelling approach," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 2, pp. 123–141, 2001.
- [9] C. Shao and X. Liu, "Estimation of Saturation Flow Rates at Signalized Intersections," *Discrete Dynamics in Nature and Society*, vol. 2012, pp. 1–9, 2012.
- [10] Y. Zhao, W. Zhang, J. Lu, W. Zhang, and Y. Ma, "Estimation of Saturation Flow Rate and Start-Up Lost Time for Signal Timing Based on Headway Distribution," *Discrete Dynamics in Nature and Society*, vol. 2015, pp. 1–7, 2015.
- [11] H. Y. Tong and W. T. Hung, "Neural network modeling of vehicle discharge headway at signalized intersection: Model descriptions and results," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 1, pp. 17–40, 2002.
- [12] 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.
- [13] Y. S. Murat and E. Gedizlioglu, "Investigation of vehicle time headways in Turkey," *Proceedings of the Institution of Civil Engineers: Transport*, vol. 160, no. 2, pp. 73–78, 2007.
- [14] X. Jin, Y. Zhang, F. Wang et al., "Departure headways at signalized intersections: a log-normal distribution model approach," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 3, pp. 318–327, 2009.
- [15] J. Tan, L. Li, Z. Li, and Y. Zhang, "A comparison of departure headway at signalized intersection produced by VISSIM and TransModeler," in *Proceedings of the Paper presented at the IEEE International Conference on Vehicular Electronics and Safety*, Beijing, China, July 2011.
- [16] S. Yin, Y. Su, C. Wang, D. Yao, L. Li, and Y. Zhang, "Comparison of vehicle departure headways in Beijing and Atlanta," *Tsinghua Science and Technology*, vol. 16, no. 3, pp. 332–336, 2011.
- [17] J. Tan, L. Li, Z. Li, and Y. Zhang, "Distribution models for start-up lost time and effective departure flow rate," *Transportation Research Part A: Policy and Practice*, vol. 51, pp. 1–11, 2013.
- [18] H. Hao and W. Ma, "Revisiting distribution model of departure headways at signalised intersections," *Transportmetrica B: Transport Dynamics*, vol. 5, no. 1, pp. 1–16, 2017.
- [19] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366, part 1, pp. 427–431, 1979.

Research Article

Optimal Bus-Bridging Service under a Metro Station Disruption

Haodong Yin,¹ Jianjun Wu ,^{1,2} Huijun Sun ,³ Yunchao Qu,¹ Xin Yang,¹ and Bo Wang⁴

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

²Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing 100044, China

³MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

⁴Beijing Municipal Commission of Transport, Beijing 100073, China

Correspondence should be addressed to Jianjun Wu; jjwu1@bjtu.edu.cn

Received 18 March 2018; Revised 10 May 2018; Accepted 22 May 2018; Published 5 July 2018

Academic Editor: Zhiyuan Liu

Copyright © 2018 Haodong Yin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A station disruption is an abnormal operational situation that the entrance or exit gates of a metro station have to be closed for a certain of time due to an unexpected incident. The passengers' travel behavioral responses to the alternative station disruption scenarios and the corresponding controlling strategies are complex and hard to capture. This can lead to the hardness of estimating the changes of the network-wide passenger demand, which is the basis of carrying out a response plan. This paper will establish a model to solve the metro station disruption problem by providing optimal additional bus-bridging services. Two main contributions are made: (1) a three-layer discrete choice behavior model is developed to analyze the dynamic passenger flow demand under station disruption; and (2) an integrated algorithm is designed to manage and control the station disruption crisis by providing additional bus-bridging services with the objective of minimizing the total travel time of affected passengers and the operating cost of bridging-buses. Besides, the multimodal transport modes, including metro, bridging-bus, shared-bike, and taxi, are considered as passengers' alternative choices in face of the station disruption. A numerical study based on the Beijing metro network shows that additional bus-bridging services can significantly eliminate the negative impact of the station disruption.

1. Introduction

Station disruption, also called station closure, is an abnormal and usually unplanned operational situation in which operators must close the entrances or exits of a metro station for various reasons, such as unexpected incidents or taking steps to avoid overcrowding [1]. In these situations, passengers cannot use the closed station as their departure station or destination stations during the closure time. Generally, disruption means a serious deviation from the planned operations in the rail transit context [2–6]. Station disruption is a typical case of metro disruption, and it can strongly affect both the service and demand of the metro system. At the service level, a station disruption means that the trains, which are planning to stop at the closed station, have to adopt other alternative routes to detour the closed station, e.g., passing the closed station without stopping. At the demand level, the

station disruption may cause significant changes in passenger flow demand at the closed stations and nearby stations.

To our knowledge, limited existing studies are available related to the topic of station disruption. Silva *et al.* [7] proposed a data-driven statistical method to determine the effect of station disruptions on the macroscopic passenger flow demand based on the smart card data from the day of the station disruption and the historical data. This approach was proved to be effective for identifying the impact of a station disruption, but it is not applicable for predicting the impact of a station disruption that does not actually occur or is about to occur. Another effective mathematical model that could capture the behavior of passengers affected by a station disruption was proposed for analyzing alternative disruption scenarios and their likely outcomes in our previous study [1]. But this model ignored the randomness of the passengers' choice behaviors and the influence of the controlling strategies

on passenger behaviors. Some other researchers studied the impacts of traffic incidents on road traffic, which could provide references for this study in metro system [8, 9].

Although we can capture the dynamic influence of the station disruption on passenger flow demand to some extent, how to minimize the severity of potential negative impacts of the station disruption by carrying the affected passengers to the destination with high service quality and low operation cost is still unsolved. Rescheduling timetables plays a limited role in solving the station disruption problem [10–18], because the station disruption defined in this paper does not interrupt the trains running process on the metro lines. But the bus-bridging service has been recognized as one of the most critical components of effective disruption responses in metro networks. Kepaptsoglou and Karlaftis [19] firstly proposed a methodological framework for planning and designing an efficient bus-bridging network. Wang et al. [20] explored the demand modeling problem for bus-bridging disruption in metro services based on the theory of compound Poisson processes and formulated it as a bulk queuing problem involving balking and reneging. Jin et al. [21] introduced a localized metro-bus integration approach that aimed at enhancing the urban transit networks resilience to disruption. Based on the previous researches, Jin et al. [22] presented an optimization-based approach of designing an optimal bus-bridging service for network disruptions by considering the commuter travel demand at the time of the disruption. But the travel demand was estimated from the historical data. In other words, the dynamic responses of passengers to the disruption were not considered.

In this paper, we attempt to manage and control the station disruption by providing additional bus-bridging services for passengers. Our contributions mainly consist of the following two parts: (1) a three-layer discrete choice behavior model is developed to analyze the dynamic passenger flow demand under station disruption; and (2) an integrated algorithm is designed to manage and control the station disruption crisis by providing additional bus-bridging services with the objective of minimizing the total travel time of affected passengers and the operating cost of bridging-buses. Besides, the multimodal transport modes, including metro, bridging-bus, shared-bike, and taxi, are considered as passengers' alternative choices in face of the station disruption.

The paper is organized as follows: Section 2 presents the detailed description of station disruption problem and the integrated optimization approach based on agent-based simulation. The case study results are reported in Section 3. Finally, conclusions are given in Section 4.

2. Bus-Bridging Service for the Station Disruption Problem

In this section, we firstly introduce the metro station disruption problem in the metro network. Then, we describe this problem with an optimization approach integrated with a three discrete choice behavior model. In this approach, an optimal bus-bridging service plan is generated by a

binary integer programming model. Besides, the passengers' response to the bus-bridging information and its randomness are also formulated in the behavior model. Finally, a solution algorithm is developed to solve the problem.

2.1. Problem Description. Consider a scenario that an emergent event causes a temporary closure of a station in the metro network. We also assume that the metro operator can accurately estimate the closure duration time and then inform the network-wide passengers immediately for simplifying the problem. Not only passengers whose departure station or destination station is the closed station are affected, but also the passengers around the closed station are potentially affected. As illustrated in Figure 1, station *B* is temporarily closed. The affected passengers planning to travel by metro have to alter to other alternative transport modes, such as taxi, shared-bike, bridging-bus, and walking.

Consequently, for passengers, they cannot board or alight the trains passing through without stopping at the closed station *B*. For the passengers with their departure stations as station *B*, they may choose other alternative stations near the closed station as their new departure stations, wait at the closed station for recovery, or take another transit mode to finish their travels. For the passengers that station *B* is their planned destinations, they have to decide whether to choose an alternative destination station or take another transit mode. The passengers make travel decisions according to their travel experiences and the released guidance information from public transit operators.

For metro operators, they must quickly carry out a responsive plan to replenish the services for the affected passengers. A common strategy is to restore the connectivity between the closed station and the rest of the metro network by running bus-bridging services. And, in station disruption problem, we consider designing the optimal direct bus-bridging routes between the closed station and the rest of the rail stations with an objective of minimizing total travel time of affected passengers and the operation cost.

In planning for responses to station disruption, challenges are to estimate the dynamic passenger demand interacting with the operational information and then to decide which stations should be connected with the closed station by bus-bridging services. Notice that when bridging-buses are operated between the closed station and nearby stations, the passengers will also decide whether to choose it to finish their travels or not. Therefore, it is a complicated decision problem: which stations among the network-wide stations should be selected to connect with the closed station with the objective of minimizing total travel time of affected passengers and the operating cost of bridging-buses, considering the dynamics of passenger behavior and demand. To describe and solve the station disruption problem, a hybrid optimization approach based on binary integer programming and passenger behavior simulation is constructed. And, in order to simplify the problem, we constrain our research to the scenario where only a single station is closed.

The basic framework of the model is shown in Figure 2. The objective function of the optimization model is minimizing the total travel time of affected passengers and the

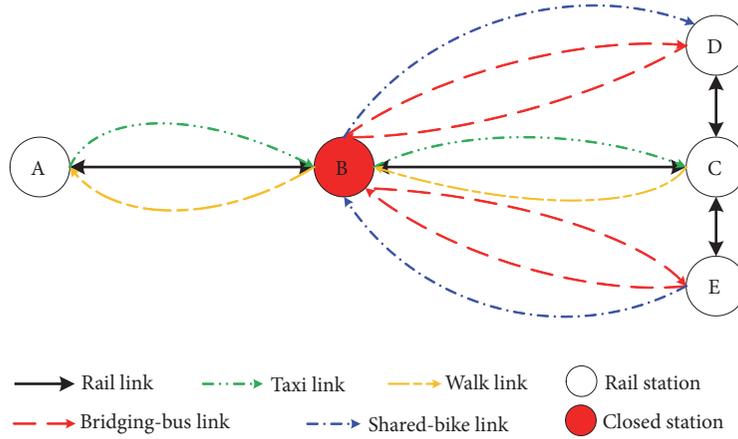


FIGURE 1: Illustrative example of station disruption problem in metro networks.

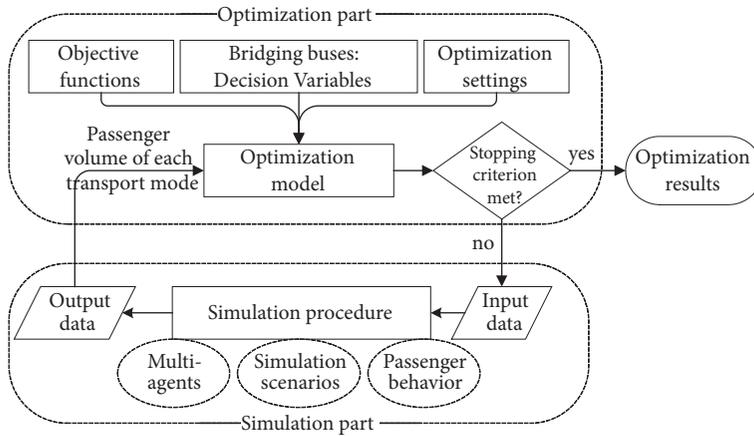


FIGURE 2: The basic framework of station disruption controlling model.

operating cost of bridging-buses. The simulation procedure is used to describe the feedback of the passenger to the bridging-buses and then to output the dynamic traffic flow for the optimization part. While the urban rail transit system involves large-scale passengers and their dynamic travel decisions under station disruption, the passenger behavior cannot be described by the precise mathematical method. Therefore, an agent-based passenger behavior simulation model is considered to describe the behavioral response of the passengers to the station disruption scenario and its corresponding bus-bridging schemes.

2.2. *Model Notations.* Main notations used in the paper are listed as follows. All variables are assumed to be integer numbers to satisfy the engineering requirements.

Parameters and Sets

$G(\mathbf{S}, \mathbf{A})$: directed graph to denote the metro network. S indicates the station set, and A indicates the set of directed arcs.

i, j, s : index of a station, $i, j, s \in S$.

a : index of an directed arc, $a \in A$.

M : the vector of transport modes, $M = \{\text{bus, metro, taxi, bike, walk}\}$, and the bus stands for the bridging-bus.

M' : the vector of transport modes except for metro, $M' = \{\text{bus, taxi, bike, walk}\}$.

m : index of transport modes, $m \in M$.

C_i^m : the carrying capacity of the transport mode m departing from the station i .

d_{ij}^m : distance of the shortest path from station i to station j by the transport mode m .

t_{ij}^m : travel time of the shortest path from station i to station j by the transport mode m .

b : index of the closed station, $b \in S$.

N : the total number of the available bridging-buses.

Decision Variables

P : set of passengers in the current metro system.

P_{ij}^t : set of passengers from station i to station j at time t , $P_{ij}^t \subset P$.

p : index of the passenger with the origin station i , destination station j , and departure time of time t , $p \in P_{ij}^t$.

α_m : the value of the travel time of urban transport mode m , which can be found in local studies in China.

β_m : the comfort penalty parameter of urban transport mode m , and there exists $\beta_{walk} > \beta_{bike} > \beta_{bus} > \beta_{taxi}$ according to our common sense.

f_{ij}^m : the fee for taking transport mode m with the shortest path from station i to station j .

$c_{ij}^{m,t}$: the generalized cost of taking transport mode m from station i to station j at time t considering the travel time and price.

$v_i^{m,t}$: the total number of passengers taking transport mode m at station i at time t .

Intermediate Variables

$x_i(t)$: indicates that whether operating bridging-buses from the closed station b to station i at time t and then back to closed station at time $t + t_{b,i}^{bus}$. 1 denotes "YES"; 0 stands for "NO."

Station Disruption Parameters

ψ_s^p : the main decision variable denotes whether station s should be chosen as the alternative origin or destination station for the passenger p , where

$$\psi_s^p = \begin{cases} 1, & \text{if station } s \text{ is selected by the passenger } p \text{ as the replaceable origin/destination station;} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

$z_{ij}^{m,p}$: the decision variable denotes whether the transport mode m should be the alternative transport

mode from station i to station j at time t for the passenger p , where

$$z_{ij}^{m,p} = \begin{cases} 1, & \text{if the transport mode } m \text{ is selected as the replaceable transport mode and } i \neq j; \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

γ_i : the constant parameter denoting whether station i is closed for the passenger p . γ_i equals 1 when station i is out of service and is 0 otherwise.

t_i^{start} : the beginning time of the disruption of station i .

t_i^{end} : the actual ending time of the disruption of station i .

$e_{ij}^{m,t}$: the extra journey time cost of metro transit from station i to station j at time t due to the station disruption:

$$e_{ij}^{m,t} = \begin{cases} t_i^{end} - t, & \text{if station } i \text{ is closed and } m = \text{metro}, t_j^{end} > t \\ t_j^{end} - t - t_{ij}^{m,t}, & \text{if station } j \text{ is closed and } m = \text{metro}, t_j^{end} > t \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

2.3. Model Formulations. The proposed integrated optimization model includes two parts. The optimization part determines which stations should be set as a bridge with the closed station while minimizing the operating cost of bridging-buses and the total travel time of affected passengers. However, the behavior model part is used to capture the passenger choice behavior response to station disruption with a three-layer discrete choice model, determining which alternative travel strategy, which alternative origin or destination, and which replaceable transport mode will be chosen by the affected passengers.

2.3.1. Optimization Model. The optimization model is used to decide which stations should be connected to the closed station around the network with a goal of minimizing the operating cost of bridging-buses and the total travel time of affected passengers. The model can be written as follows:

$$\begin{aligned} \min \quad & Z = 2\mu \sum_{i \in S, i \neq b} (x_i(t) * d_{b,i}^{bus}) + \gamma \sum_{i \in S} g(x_i) \\ \text{s.t.} \quad & g(x_i) \end{aligned} \quad (4)$$

$$= \sum_{p \in P_{b,i}^t} \sum_{s \in S} \psi_s^p \left(\sum_{m \in M'} z_{b,s}^{m,p} t_{b,s}^m + t_{s,i}^{metro} + e_{s,i}^{metro,t} \right) + \sum_{p \in P_{i,b}^t} \sum_{s \in S} \psi_s^p \left(t_{i,s}^{metro} + e_{i,s}^{metro,t} + \sum_{m \in M'} z_{s,b}^{m,p} t_{s,b}^m \right) \quad (5)$$

$$\sum_{s \in S} \psi_s^p = 1, \quad \forall p \quad (6)$$

$$\sum_{m \in M'} z_{ij}^{m,p} \leq 1, \quad \forall p, i \neq j \in S \quad (7)$$

$$\sum_{i \in S} \sum_{p \in P_{b,i}^t} \sum_{s \in S} \psi_s^p z_{b,s}^{m,p} \leq C_b^m, \quad \forall m \in M' \quad (8)$$

$$\sum_{i \in S} \sum_{p \in P_{i,b}^t} \psi_s^p z_{s,b}^{m,p} \leq C_s^m, \quad \forall m \in M' \quad (9)$$

$$x_i(t) = 0, 1, \quad \forall i \neq b \in S \quad (10)$$

As indicated in (4), the objective function contains two parts: The first part is the operating cost of bridging-buses, running between the disrupted station and other stations. The second part is the total travel time of all affected passengers travelling with various transport modes.

Equation (5) defines the function to calculate the total travel time of all affected passengers. In the function, the affected passengers are defined as two types: one is the passenger whose planned origin station is closed, and the other is the passenger whose planned destination is closed.

The travel time of two types of passengers is calculated differently.

Equations (6)-(7) indicate the constraints of the passenger behavioral choices. Equation (6) shows that an alternative origin or destination station will be chosen for a particular passenger p . Constraint (7) indicates that one new transport mode will bridge the invalid origin station and the new replaceable station, or the new destination station to the invalid destination station.

Equations (8)-(9) indicate the capacity constraints of multimodal transport resources. Constraint (8) denotes the capacity constraints of the transport mode m at the closed station b . Constraint (9) denotes the capacity constraints of the transport mode m at the closed station s .

Equation (10) shows the decision variables are binary.

2.3.2. Passenger Behavior Model. The lower model is to capture the behavioral responses to the station disruption and its corresponding bus-bridging strategies and considering the randomness of the choice behavior on alternative stations or transport modes, e.g., bridging-buses, taxis, metro, shared-bicycles, and walking.

Generate the Optimal Travel Decision. For a passenger $p \in P_{O,D}^t$, the objective of passenger behavior model is to minimize the generalized travel cost of making a new travel decision for the station disruption. The decision variables are ψ_s^p and $z_{ij}^{m,p}$.

$$\text{Objective min} \quad \sum_{i \in \{O,D\}} \gamma_i^p J_i^p \quad (11)$$

Subject to

$$J_O^p = \sum_{s \in S} \psi_s^p \left(\sum_{m \in M'} z_{ij}^{m,p} c_{O,s}^{m,t} + c_{s,D}^{metro,t} \right), \quad \forall t, p, \text{ if } O \text{ is out of service} \quad (12)$$

$$J_D^p = \sum_{s \in S} \psi_s^p \left(c_{O,s}^{metro,t} + \sum_{m \in M'} z_{ij}^{m,p} c_{s,D}^{m,t} \right), \quad \forall t, p, \text{ if } D \text{ is out of service} \quad (13)$$

$$\gamma_O^p + \gamma_D^p = 1, \quad \forall p \quad (14)$$

$$\sum_{m \in M'} z_{ij}^{m,p} = 1, \quad \forall p, i \neq j \quad (15)$$

$$\sum_{s \in S} \psi_s^p = 1, \quad \forall p \quad (16)$$

$$c_{ij}^{m,t} = \begin{cases} W, & \text{if } v_i^{m,t} > C_i^m; \\ \beta_m (\alpha_m (t_{ij}^m + e_{ij}^{m,t}) + f_{ij}^m) + (1 - x_i(t)) W, & \text{if } m = 'bus', j = b; \\ \beta_m (\alpha_m (t_{ij}^m + e_{ij}^{m,t}) + f_{ij}^m) + (1 - x_j(t)) W, & \text{if } m = 'bus', i = b; \\ \beta_m (\alpha_m (t_{ij}^m + e_{ij}^{m,t}) + f_{ij}^m), & \text{elsewhere.} \end{cases} \quad \forall t, m, i, j \quad (17)$$

$$t_i^{start} \leq t \leq t_i^{end}, \quad \forall i = O \text{ or } D \quad (18)$$

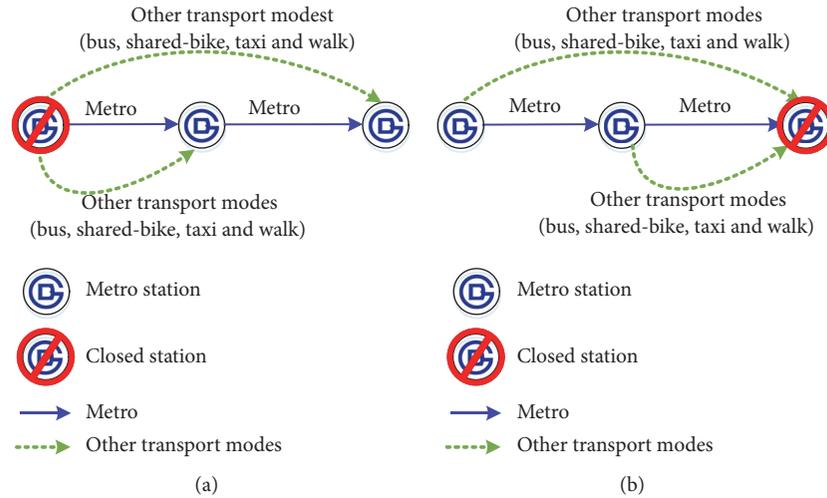


FIGURE 3: The two station disruption situations and likely choice behavior for a passenger. (a) shows the situation where the passenger's origin station is closed. (b) shows the situation that the passenger's destination station is closed.

Objective function (11) minimizes the generalized travel cost. Equations (12)–(14) indicate the basic constraints of the station disruption situation for a particular passenger, as illustrated in Figure 3. Constraint (12) indicates the generalized travel cost of making a new travel decision if the passenger's origin station is closed. Constraint (13) denotes the generalized travel cost of making a new travel decision if the passenger's destination station is closed. Constraint (14) indicates that only one case occurs for a particular passenger, which means that either the origin station ($\gamma_O^p = 1$, $\gamma_D^p = 0$, i.e., $b = O$) or the destination station ($\gamma_O^p = 0$, $\gamma_D^p = 1$, i.e., $b = D$) is closed.

Equations (15)–(16) denote the limits of the decision variables. Equation (15) denotes that another transport mode will bridge the closed origin station and the new departure station, or the new destination station and the closed destination station except the metro. Equation (16) denotes that an alternative station will be chosen to replace the closed one.

Equation (17) indicates the generalized cost of taking transport mode m from station i to station j at time t considering the travel time, price, and comfort penalty. W is defined as a very large constant. When $x_i(t) = 0$, it means that the generalized travel cost will be very large if the passenger p still chooses the bridging-bus as his alternative transport mode. It also indicates that only when bridging-buses are operated between stations and the closed station b , the bridging-buses can be available for passengers ($x_i(t) = 1$). It can be seen that the passenger's travel behavior is affected by the upper decision variables of controlling measures.

Equation (18) denotes that the time interval for a particular passenger to make a decision is only within the period of the station disruption. In other words, the model is applicable when the affected passenger has just been faced with the station disruption.

Three-Layer Discrete Choice Behavior Model for Alternative Travel Strategies. In a previous section, a behavior optimization model is proposed to generate an optimal travel strategy with a minimum generalized travel cost for the case of a station disruption. But, in an actual disruption scenario, the passengers' travel experience partially or completely turns into failure, and the necessary information for decision-making is usually lacking. Consequently, passengers cannot make a rational or optimal travel decision. Therefore, the randomness of passenger travel behavior should be considered in the newly proposed model.

Based on the stochastic utility theory, a three-layer nested-logit model for passengers' choice behavior under station disruption is constructed, which is illustrated in Figure 4.

In the proposed three-layer nested-logit model, the first layer is to make a trip decision among quitting the metro journey (labeled as the event y_1), waiting for the recovery of station service (labeled as event y_2) and taking an alternative metro station (labeled as event y_3); the second layer is to make travel decisions among the alternative stations set for the passengers taking an alternative metro station, or choices among various transport modes to continue the journey; the third layer is to make choices among available transportation modes to bridge the new alternative metro station and the planned origin or destination station. The multimodal transport resources contain bridging-bus, taxi, shared bicycle, and walking, which are labeled as $m1$, $m2$, $m3$, and $m4$, respectively. Accordingly, the general cost of each behavior is c_1 , c_2 , and c_3 , respectively, for events y_1 , y_2 , and y_3 . Take a passenger $p \in P_{O,D}^t$ as an example, and we can have these remarks.

(a) For the passenger who gives up the metro journey, the choice behavior is that the passenger will take the planned destination as his new departure station in the situation of

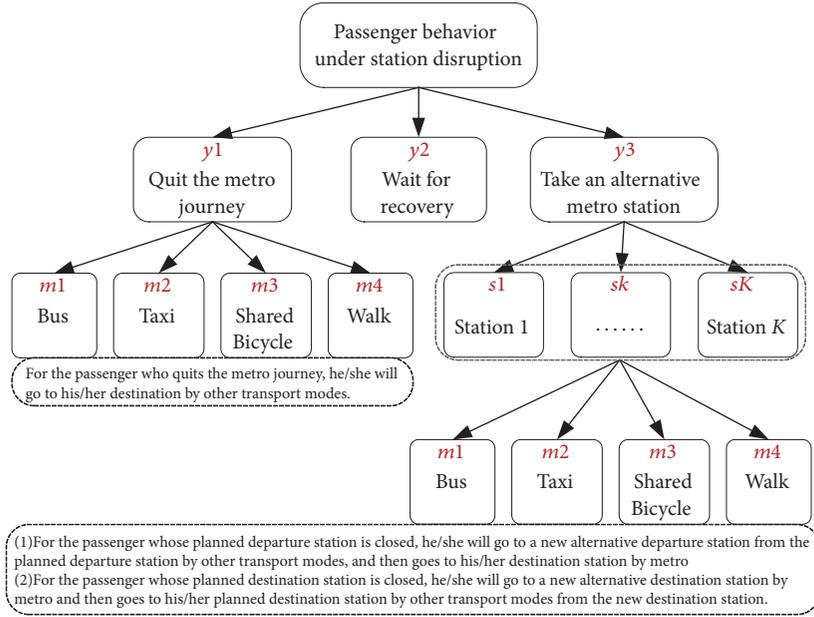


FIGURE 4: The nested-logit model for passenger behavioral choices under station disruption.

$\gamma_O^p = 0$ and $\gamma_D^p = 1$, or take the planned origin station as his new destination in the situation of $\gamma_O^p = 1$ and $\gamma_D^p = 0$. Then, the probability of choosing the model m can be calculated as

$$P(m | y_1) = \frac{\exp(-\theta c_{OD}^{m,t} / \min_{l \in M'} \{c_{OD}^{l,t}\} + b)}{\sum_{m \in M'} \exp(-\theta c_{OD}^{m,t} / \min_{l \in M'} \{c_{OD}^{l,t}\} + b)} \quad (19)$$

$\forall m \in M'$

So c_1 can be calculated based on (12) ~ (14):

$$c_1 = \sum_{m \in M'} P(m | y_1) c_{OD}^{m,t} \quad (20)$$

(b) For the passenger who waits for the recovery of station service, the choice behavior is that the passenger will take the planned origin station as his new departure station. Once the destination is closed, it is hard to decide which station should be chosen to wait at. So waiting for the recovery of station service can be one option only in the case that one's planned departure station is closed. So c_2 can be calculated based on (12) ~ (14):

$$c_2 = c_{OD}^{rail} \quad (21)$$

(c) For the passenger who takes an alternative metro station, the choice behavior is that the passenger will take an alternative station as his new departure station or destination station. If the station k is selected, the generalized travel cost of the passenger choosing the transport mode m can be calculated as

$$c_3^{m,k,t} = \begin{cases} c_{O,k}^m + c_{k,D}^{metro}, & \text{if } O \text{ is closed;} \\ c_{O,k}^{metro} + c_{k,D}^m, & \text{if } D \text{ is closed.} \end{cases} \quad (22)$$

Firstly, in this situation, the probability of choosing the model m can be calculated as

$$P(m | k, y_3) = \frac{\exp(-\theta c_3^{m,k,t} / \min_{l \in M'} \{c_3^{l,k,t}\} + b)}{\sum_{r \in M'} \exp(-\theta c_3^{r,k,t} / \min_{l \in M'} \{c_3^{l,k,t}\} + b)} \quad (23)$$

We label c_3^k as the generalized travel cost of the passenger p who chooses the station k as his new alternative departure or destination station:

$$c_3^k = \sum_{m \in M'} P(m | k, y_3) c_3^{m,k,t} \quad (24)$$

Then, the probability of choosing the station k can be calculated as

$$P(k | y_3) = \frac{\exp(-\theta c_2^k / \min_{l=1 \dots K} \{c_3^l\} + b_1)}{\sum_i^K \exp(-\theta c_3^i / \min_{l=1 \dots K} \{c_3^l\} + b)} \quad (25)$$

$\forall k = 1 \dots K$

$$c_3 = \sum_k P(k | x_2) c_2^k \quad (26)$$

The probability of each passenger behavior decision is

$$P(y_j) = \frac{\exp(-\theta c_j / \min_{l=1,2,3} \{c_l\} + b_2)}{\sum_{i=1}^3 \exp(-\theta c_i / \min_{l=1,2,3} \{c_l\} + b)} \quad (27)$$

$\forall j = 1, 2, 3$

The joint probability of passenger choosing a station k as an alternative station can be calculated:

$$P(k, y_3) = P(k | y_3) P(y_3) \quad (28)$$

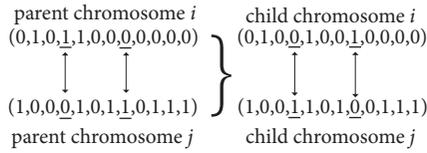


FIGURE 5: The crossover operator.

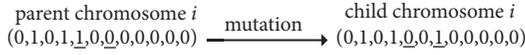


FIGURE 6: The mutation operator.

Furthermore, the joint probability of choosing transport mode m can be

$$P(m, y_1) = P(m | y_1) P(y_1) \quad (29)$$

$$P(m, k, y_3) = P(m | k, y_3) P(k, y_3) \quad (30)$$

2.4. An Integrated Solution Algorithm

2.4.1. Genetic Algorithm. The decision variable in the upper model is $x_i(t)$, indicating whether the bridging-bus should be operated between the closed station and the station $i \in S$ at the time interval t . 1 denotes “YES” and 0 stands for “NO.” Therefore, in the genetic algorithm, the traditional binary coding method can be used to design a feasible solution as the corresponding gene sequence or chromosome. A chromosome is represented by a vector as indicated in (31). Besides, the population is represented by H in (32).

$$H_i = ({}_1x_0, \dots, {}_1x_i, \dots, {}_1x_n) \quad (31)$$

$$H = \{H_1, H_2, \dots, H_L\} \quad (32)$$

Generally, the genetic algorithm consists of three important operators: the crossover operator, the mutation operator, and the selection operator. We assume that the chromosome intersects with a certain probability and randomly selects the intersection with a certain crossover probability ξ . In the algorithm, the crossover method is given as shown in Figure 5. Since the chromosome is composed of a specific gene fragment which cannot be divided only the same position of the gene fragment will be crossed.

The mutation operator can help avoid local optimal solutions. In this paper, we assume that each chromosome will be a certain probability of mutation operation ζ and randomly select the mutation position, as shown in Figure 6.

The selection operator is to ensure the best individuals can survive in the next generation. And it is designed based on the fitness function, defined as $f(n) = 1/Z$, where Z is defined by formula (4). The more the fitness is, the more likely the best individual can survive. We adopt the Roulette method to determine whether each individual can survive in the next generation or not.

2.4.2. Solution Algorithm of Passenger Behavior Model. The general simulation procedure can be attributed to Yin et al.

[1]. A new behavior model solver in the simulation algorithm is developed for the three-layer discrete choice behavior model, which is summarized as shown in Algorithm 1:

2.4.3. The General Solution Algorithm. In this paper, a genetic algorithm and a simulation based method are integrated to solve the proposed model. The genetic algorithm is used to solve the binary integer programming model and the simulation procedure is to generate different types of the passenger flow demand. And the basic simulation procedure is attributed to Yin et al. [1]. The general algorithm framework for solving the problem of station disruption controlling optimization is described as follows.

Step 1 (initialize). Initialize the parameters and set the maximum number of iterations to G , and the population size is L . Initialize the disruption's time duration as $[1, n]$. Set the current time interval $t = 1$.

Step 2. Set an initial set of bus-bridging scheme or initial population. Set the current generation $g = 0$.

Step 3. Solve the lower model. For a particular bus-bridging scheme, the simulation model and algorithm are called to capture the passenger travel behavior and each passengers' travel time. And calculate Z .

Step 4. Solve the upper model. According to each passengers' travel time, the fitness of each chromosome in the initial population can be calculated and through the following operations, to generate the upper optimization of the bus-bridging scheme strategy x_i .

Step 4.1. Mutation operation: using the mutation operator to create a new chromosome.

Step 4.2. Cross operation: cross the operation on the L chromosome.

Step 4.3. Selection operation: using the selection operator to ensure that the optimal individual genetic to the next generation.

Step 5. Verify the end condition 1. If the number of iterations is greater than G , then the optimal solution of the model will be obtained; otherwise, $g = g + 1$, and the iteration will return to Step 3.

Step 6. Verify the end condition 2. If $t > n$, end; otherwise, make $t = t + 1$, and return to Step 2.

3. Numerical Example

In this section, we present the application of suggested model and the solution algorithm to an example urban metro network based on real and virtual data. Figure 7 illustrates the example station disruption scenario, which contains 2 directional lines and 2 transfer stations extracted from Beijing Subway network, which is one of the busiest metro systems [23]. Figure 8 illustrates the practical train timetables. We

Initialization: initialize the corresponding parameters of discrete choice behavior model, and K .

Procedure:

- (1) **if** $\gamma_O^p == 1$ **then**
- (2) Define a temporary set $List_k$;
- (3) **for** $s \in S$ **do**
- (4) **if** $s == O$ **then**
- (5) Calculate the c_2 based on the Eq. (21).
- (6) **else if** $s == D$ **then**
- (7) Calculate the c_1 based on the Eqs. (19)~(20).
- (8) **else**
- (9) Calculate the c_3^s based on the Eqs. (22)~(24).
- (10) if the number of elements in $List_k$ is not exceeding K , add the c_3^s into the set $List_K$ and sort the elements from smallest to largest;
- (11) if the number of elements in $List_k$ equals K , and the last value of $List_k$ is larger than c_3^s , remove the last value of $List_k$ and add c_3^s to the $List_k$.
- (12) **end if.**
- end for.**
- (13) Calculate the c_3 based on the $c_3^s \in List_k$ using the Eqs. (25)~(26).
- (14) Then, calculate $P(y_j)$ based on Eqs. (27).
- (15) Calculate the joint probability $P(m, y_1), P(k, y_3), P(m, k, y_3)$ based on Eqs. (28)~(30).
- (16) Using the Roulette method to generate an alternative travel decision:
- (17) (1) Construct a cumulative probability set R based on $P(m, y_1), P(y_2), P(m, k, y_3)$, where $m=1, 2, 3, 4$, and $k = 1, 2, \dots, K$.
- (18)
$$R[m-1] = \sum_{l=1}^m P(l, y_1), \quad m = 1, 2, 3, 4$$
- (19)
$$R[4k+m-1] = \sum_{l=1}^4 P(l, y_1) + \sum_{u=1}^{k-1} \sum_{v=1}^4 P(v, u, y_3) + \sum_{v=1}^m P(v, k, y_3), \quad m = 1, 2, 3, 4, k = 1, 2, \dots, K.$$
- (20)
$$R[4(K+1)] = \sum_{l=1}^4 P(l, y_1) + \sum_{u=1}^K \sum_{v=1}^4 P(v, u, y_3) + P(y_2) = 1$$
- (21) (2) Generate a random variable r .
- (22) (3) Determine the behavior choices:
- (23) if $r \leq R[0]$, passenger will give up the metro journey and take a bus instead to the destination and $v_O^{bus,t} + +$;
- (24) if $R[m-1] < r \leq R[m]$, passenger will give up the metro journey and take the transport mode m instead to the destination, $v_O^{m,t} + +$;
- (25) if $R[4k+m-1] < r \leq R[4k+m]$, passenger will take an alternative station k as his new origin station of metro journey, and he will take the transport mode m to the new origin station from the planned origin station, the entry time at new origin is the original entry time added by the time consumption of the transport mode m , $v_O^{m,t} + +$;
- (26) if $R[4K+3] < r \leq 1$, passenger will wait at the planned origin station.
- (27) **else if** $\gamma_D^p == 1$ **then**
- (28) Define a temporary set $List_k$, and we set the O as the passengers' current station at the current time. If passenger p is now in train, we set the O as the next station the train will stop at.
- (29) Judge whether the travel time from the current station to the destination is larger than the closure duration? if yes, the passenger p need not to change his travel behavior. If not, go on executing the following part.
- (30) **for** $s \in S$ **do**
- (31) **if** $s == O$ **then**
- (32) Calculate the c_1 based on the Eqs. (19)~(20).
- (33) **else if** $s == D$ **then**
- (34) Set $c_2 = W$, where W is a very big constant.
- (35) **else**
- (36) Calculate the c_3^s based on the Eqs. (22)~(24).
- (37) if the number of elements in $List_k$ is not exceeding K , add the c_3^s into the set $List_K$ and sort the elements from smallest to largest;
- (38) if the number of elements in $List_k$ equals K , and the last value of $List_k$ is larger than c_3^s , remove the last value of $List_k$ and add c_3^s to the $List_k$.
- (39) **end if.**
- (40) **end for.**
- (41) Calculate the c_3 based on the $c_3^s \in List_k$ using the Eqs. (25)~(26).
- (42) Then, calculate $P(y_j)$ based on Eq. (27).
- (43) Calculate the joint probability $P(m, y_1), P(k, y_3), P(m, k, y_3)$ based on Eqs. (28)~(30).

- (44) Using the Roulette method to generate an alternative travel decision:
(45) (1) Construct a cumulative probability set R based on $P(m, y_1)$, $P(y_2)$, $P(m, k, y_3)$, where $m=1, 2, 3, 4$, and $k = 1, 2, \dots, K$.
(46) $R[m - 1] = \sum_{l=1}^m P(l, y_1)$, $m=1, 2, 3, 4$
(47) $R[4k + m - 1] = \sum_{l=1}^4 P(l, y_1) + \sum_{u=1}^{k-1} \sum_{v=1}^4 P(v, u, y_3) + \sum_{v=1}^m P(v, k, y_3)$, $m=1, 2, 3, 4$, $k = 1, 2, \dots, K$.
(48) (2) Generate a random variable r .
(49) (3) Determine the behavior choices:
(50) if $r \leq R[0]$, passenger will give up the metro journey and take a bus instead to the destination from current station, $v_O^{\text{bus},t} + +$;
(51) if $R[m - 1] < r \leq R[m]$, passenger will give up the metro journey and take the transport mode m instead to the destination from current station, $v_O^{m,t} + +$;
(52) if $R[4k + m - 1] < r \leq R[4k + m]$, passenger will take an alternative station k as his new destination station of metro journey, and he will take the transport mode m to the planned destination station from the new destination station, $v_k^{m,t} + +$;
(53) **end if.**

ALGORITHM 1: Solution algorithm of the passenger behavior model.

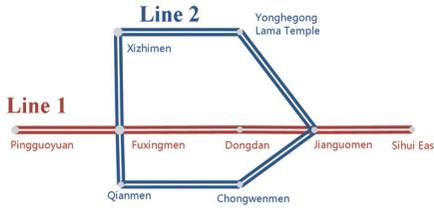


FIGURE 7: An example of a metro network extracted from Beijing metro network.

assume that the Xizhimen station is closed. The cost of taking alternative transport modes is given in Table 1.

In order to compute and test the model more easily, we assume that each OD volume shares the same distribution, as shown in Figure 9. Different values of V can be investigated for different scales of passenger demand. And other values of main parameters are defined in Table 2 for simplification.

$$f^{od}(t) = \begin{cases} \frac{V}{12} \times t, & t \in [0, 12) \\ -\frac{V}{16} \times t + \frac{7V}{4}, & t \in [12, 28) \end{cases} \quad (33)$$

where t denotes the t th time interval of 15 minutes and V indicates the maximum passenger demand of each two different stations per time interval, indicated in Figure 9.

3.1. Passenger Flow Demand Analysis under a Station Disruption

3.1.1. Analysis of Individual Behavioral Choices

Passengers' Behavior Choices under Different the Closure Duration. For a particular OD from the Xizhimen station to the Pingguoyuan station, Figure 10 shows that, with increasing closure duration, the passengers will more likely alter their planned origin stations or give up their metro journeys, and fewer passengers will wait for its recovery of station service.

Alternatives of Transport Modes. Figure 11 shows the passengers' choices in multimodal transports. With increasing the bridging distance up to 6 km, more passengers prefer to choose the taxi as their alternative transport mode.

Alternatives of Origin Stations. Figure 12 shows the top stations that passengers prefer to choose as their alternative origins when their planned origin station Xizhimen station is closed for different durations. The Fuxingmen, which is near the Xizhimen station, is always chosen by most passengers. The result also shows that fewer passengers will wait at the closed station and more passengers will give up their planned origin stations to nearby stations if the closure duration becomes larger.

3.1.2. Passenger Flow Demand at Affected Stations

Passenger Flow Dynamics at the Closed Station. In a 60 minutes' station disruption, the number of entries at the Xizhimen station is changing over time, as is shown in Figure 13. It also shows that when the disruption comes to an end, the passengers waiting at closed station will gradually enter the closed station, which leads to more entries than normal. And, consequently, it takes longer than the closure duration itself for the entries to recover to a normal level, as shown in Figure 13.

Passenger Flow Dynamics at the Nearby Stations. The disruption of Xizhimen station results in a significant increase in the number of entries at the nearby stations, especially at the Fuxingmen station, as is shown in Figure 14.

3.2. Optimization Results.

Optimization Results under Different Disruption Durations. As is indicated in Figure 15, a set of disruption scenarios are numerically computed for a determined OD volume of $V=240$, and the duration of station disruption ranges from 0 to 120 minutes. The results show that when the duration

TABLE 1: Travel time and the price of taking alternative transport modes.

origin	Destination	Travel time by walk(s)	Price of walk (Yuan)	Travel time by bike(s)	Price of bike (Yuan)	Travel time by bus(s)	Price of bus (Yuan)	Travel time by taxi(s)	Price of taxi (Yuan)
Xizhimen	Pinguoyuan	12000	0	6239	4	4779	5	2913	47
Pinguoyuan	Xizhimen	11741	0	5982	4	4782	5	3045	52
Xizhimen	Fuxingmen	2749	0	1356	1	2031	2	855	18
Fuxingmen	Xizhimen	2637	0	1366	1	1673	2	815	17
Xizhimen	Dongdan	6153	0	3161	2	3591	2	1777	28
Dongdan	Xizhimen	6134	0	3166	2	3334	2	1300	28
Xizhimen	Jianguomen	7128	0	4000	3	4799	4	1922	31
Jianguomen	Xizhimen	9251	0	3647	3	3550	3	1484	34
Xizhimen	Sihui East	66667	0	6167	4	5520	5	1740	65
Sihui East	Xizhimen	66667	0	6024	4	5520	5	1740	65
Xizhimen	Qianmen	5955	0	2801	2	3925	2	1739	26
Qianmen	Xizhimen	6571	0	2849	2	4001	2	1257	27
Xizhimen	Chongwenmen	6828	0	3385	2	4200	3	1991	29
Chongwenmen	Xizhimen	6817	0	3490	2	3679	3	1598	31
Xizhimen	Yonghegong Lama Temple	4155	0	1981	2	2199	2	1208	22
Yonghegong Lama Temple	Xizhimen	5059	0	1993	2	2258	2	983	21

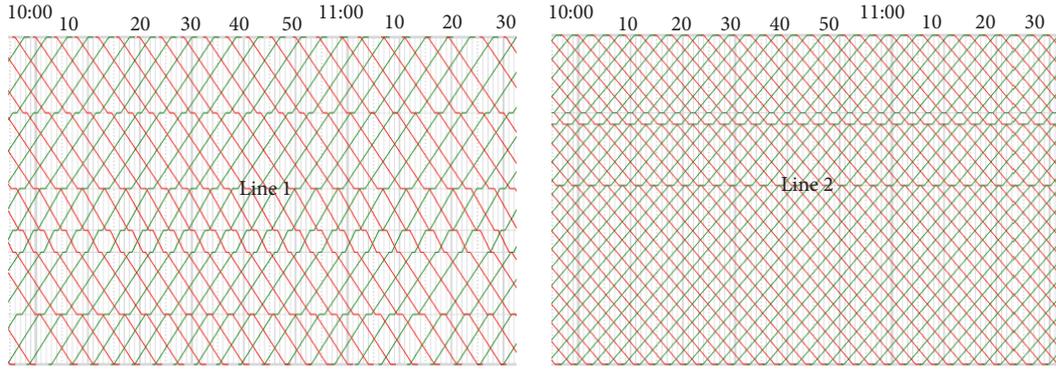


FIGURE 8: The partial train diagram of each metro line. The horizontal axis denotes the time. For example, the “10” after the “10:00” indicates the time of 10:10, and so on.

TABLE 2: Values of some parameters in the numerical study.

Parameters	Definitions	Values
μ	the parameter of transferring the buses' running distance to the standardized operating cost	100
γ	the parameter of transferring the passengers' travel time to the standardized travel cost	1
θ	the magnification parameter of the generalized cost in the logit model	3
b	a random constant associated with the generalized cost in the logit model	0
α_m	the value of the travel time of urban transport mode m , which can be found in local studies in China	60
β_{walk}	the comfort penalty parameter of walk	3
β_{bike}	the comfort penalty parameter of bike	1.8
β_{bus}	the comfort penalty parameter of bus	1.5
β_{taxi}	the comfort penalty parameter of taxi	1.2
ξ	crossover probability in genetic algorithm	0.2
ζ	mutation probability in genetic algorithm	0.2
G	the upper limit of the number of iterations in the genetic algorithm	100
L	population size in genetic algorithm	200

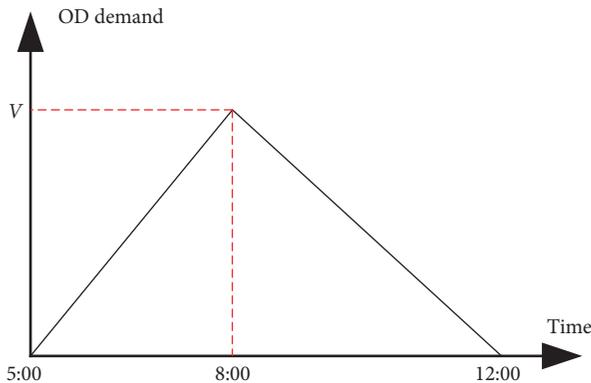


FIGURE 9: The OD passenger flow demand over time.

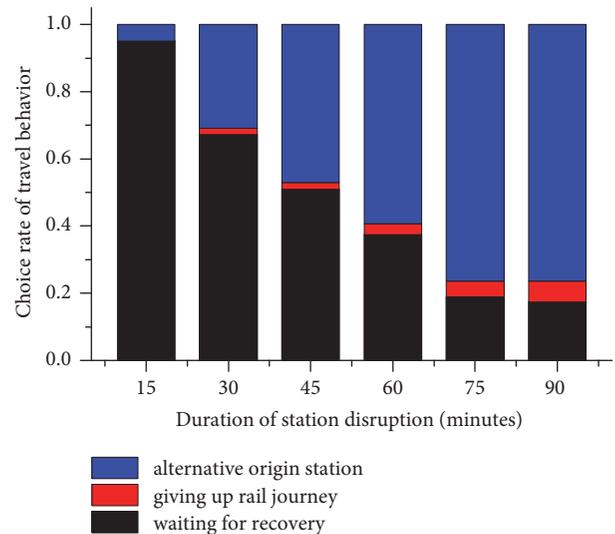


FIGURE 10: The relationship between the duration of station disruption and the passenger behavioral choices.

of station disruption becomes larger, the total travel time will be more significantly reduced if the bridging-buses are operated. The improvement can reach up to 49.38% under a station disruption of 120 minutes. That is because when

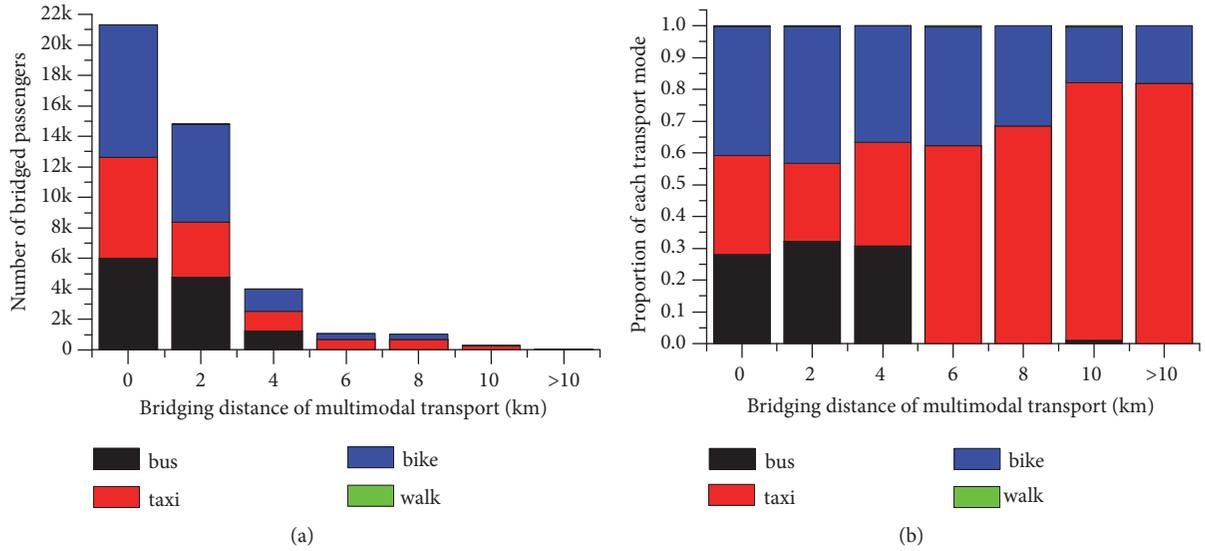


FIGURE 11: Alternatives of transport modes. (a) shows the numbers of bridging passengers are decreasing sharply with the increases in bridging distance; (b) shows the proportion of passengers choosing the different transport mode with the changes of bridging distance.

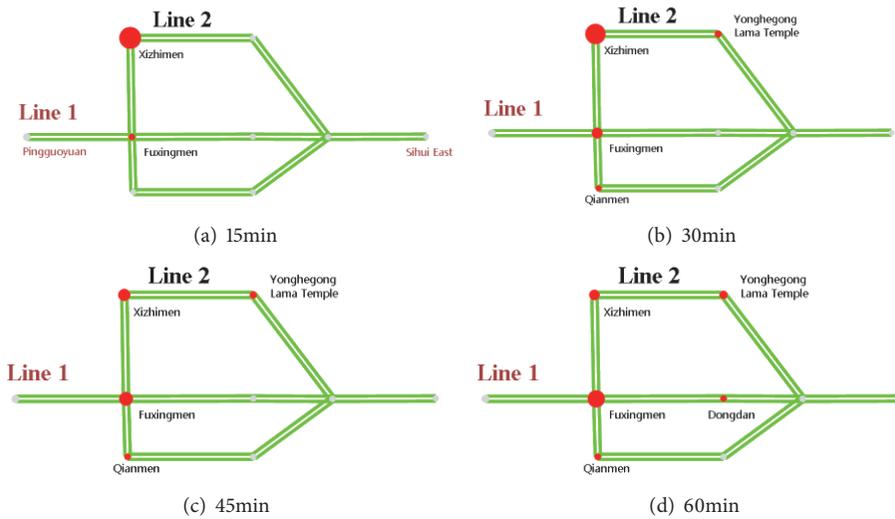


FIGURE 12: The replaceable departure stations passengers prefer to choose when Xizhimen is closed for different closure durations. The red circles represent the alternative origin stations, including the closed station. The size of each circle represents the number of passengers who choose that station, and a larger radius indicates relatively more passengers. (a) shows the duration of 15 minutes; (b) shows the duration of 30 minutes; (c) shows the duration of 45 minutes; (d) shows the duration of 60 minutes.

the station is closed for a longer time, more passengers will choose the bridging-buses or other transport modes instead of waiting for recovery of the station service, which leads to a better controlling effect. It also indicates that, in the minor disruption scenario, there is no need to operate the bridging-buses.

The blue line in Figure 15 also shows that our developed solution algorithm is time-expensive. It costs up to 305 minutes in an only 120 minutes' station disruption scenario in the small numerical study, which seems not to be efficient enough for large-scale network cases or on-line emergent

situations. That is because our proposed model must calculate all behavioral responses to the station disruption for every affected passenger in each iteration in the solution algorithm.

Optimization Results under Different Passenger Flow. As is indicated in Figure 16, another set of disruption scenarios are numerically computed for a disruption duration of 60 minutes, and the OD volume of V ranges from 60 to 420. The results show that, with increasing the passenger flow demand, more total travel time will be significantly reduced if the bridging-buses are operated.

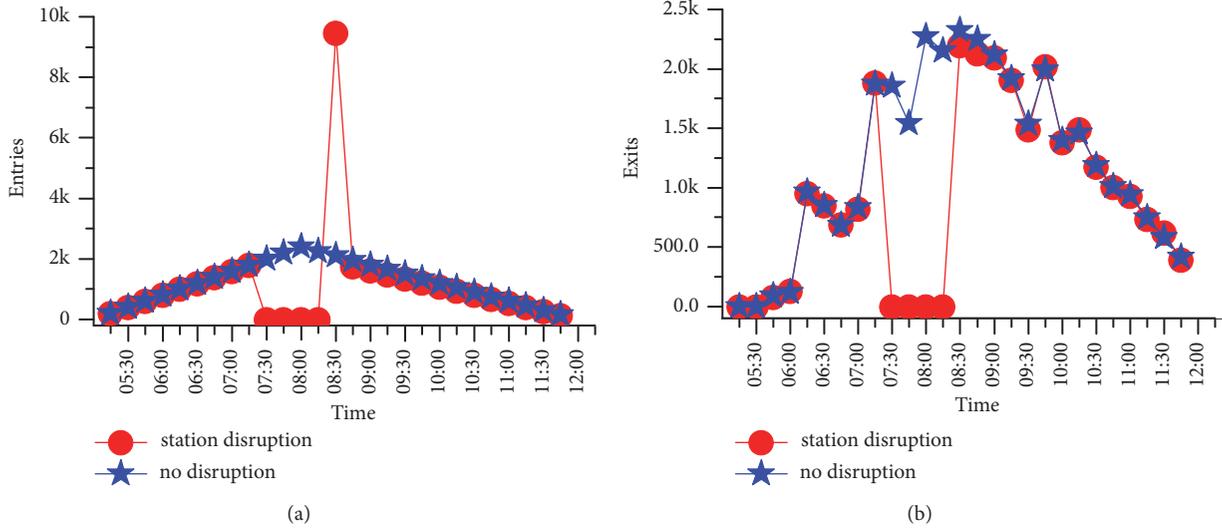


FIGURE 13: The number of entries and exits at the closed station Xizhimen. (a) shows entries of Xizhimen over time; (b) shows the exits of Xizhimen over time.

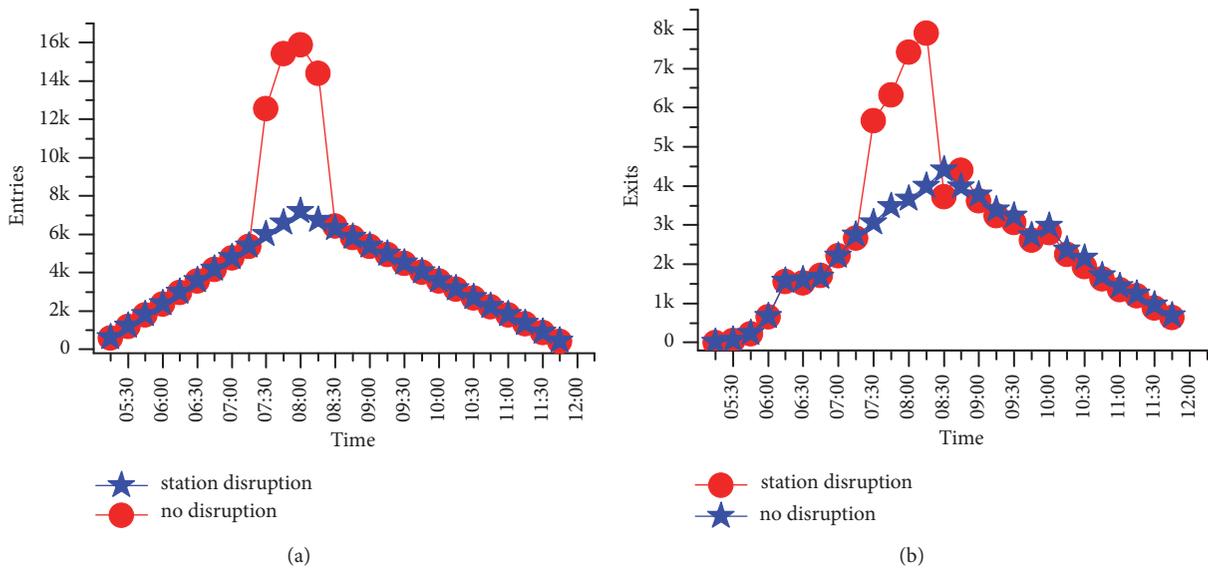


FIGURE 14: The number of entries and exits at the affected station-Fuxingmen. (a) shows entries of Fuxingmen over time; (b) shows the exits of Fuxingmen over time.

4. Conclusion

An integrated optimization approach based on behavior model is proposed to manage and control the disruption crisis by providing additional bus-bridging services with the objective of minimizing the total travel time of affected passengers and the operating cost of bridging-buses. Among them, a three-layer discrete choice behavior model is proposed to analyze the passenger flow demand under station disruption. Furthermore, the multimodal transport modes, including metro, bridging-bus, shared-bike, and taxi, are considered as passengers' alternative choices in face of the station disruption. A numerical study based on Beijing metro

network is conducted finally. Some conclusions can be drawn as follows.

(a) The effects of a station disruption with different durations on passenger flow demand are analyzed, and the results show that, with increasing closure duration, more passengers will alter their planned origin stations or quit their metro journey, and fewer passengers will wait at the closed station for its recovery. And, with increasing the bridging distance up to 6 km, more passengers prefer to choose the taxi as their alternative transport mode.

(b) The optimization results show that additional bus-bridging services can eliminate the station disruption impact significantly, especially with the increase of closure duration

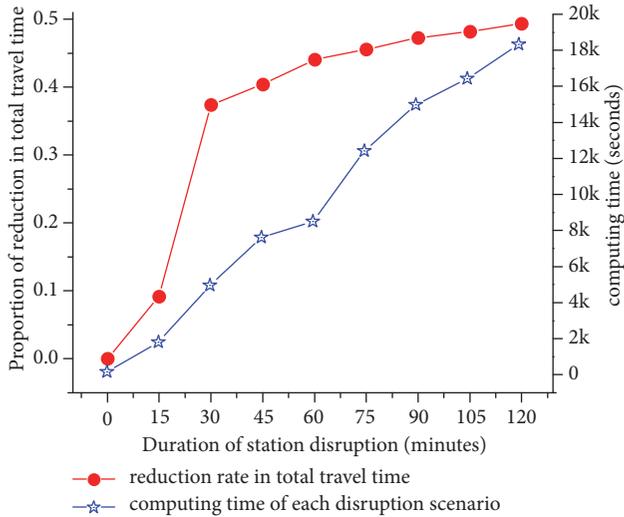


FIGURE 15: The optimization results under different disruption durations.

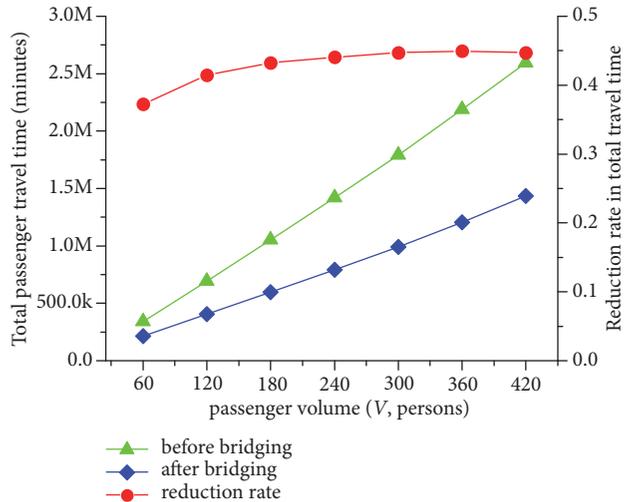


FIGURE 16: The optimization results under different passenger flow demand.

and the passenger flow demand. The improvement can reach up to 49.38% under a station disruption of 120 minutes and a determined OD volume of $V=240$ in our numerical study.

(c) Our developed solution algorithm is time-expensive and seems not to be efficient and applicable enough for large-scale network cases or real-time emergent situations. A more efficient algorithm should be developed in our future studies.

Data Availability

All data generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by NSFC (nos. 71771018, 71621001, 71701013, and 71601017), the Fundamental Research Funds for the Central Universities (no. 2018JBM074), Research Foundation of State Key Laboratory of Rail Traffic Control and Safety (no. RCS2018ZT004), and Technology Foundation of Beijing Jiaotong University (no. 2018RC010). The authors thank these agencies.

References

- [1] H. Yin, B. Han, D. Li, J. Wu, and H. Sun, "Modeling and simulating passenger behavior for a station closure in a rail transit network," *PLoS ONE*, vol. 11, no. 12, Article ID e0167126, 2016.
- [2] L. Cadarso, Á. Marín, and G. Maróti, "Recovery of disruptions in rapid transit networks," *Transportation Research Part E: Logistics and Transportation Review*, vol. 53, no. 1, pp. 15–33, 2013.
- [3] X. He and H. X. Liu, "Modeling the day-to-day traffic evolution process after an unexpected network disruption," *Transportation Research Part B: Methodological*, vol. 46, no. 1, pp. 50–71, 2012.
- [4] J. Jespersen-Groth, D. Potthoff, J. Clausen et al., "Disruption management in passenger railway transportation," in *Robust and Online Large-Scale Optimization*, R. K. Ahuja, R. H. Möhring, and C. D. Zaroliagis, Eds., pp. 399–421, Springer, Berlin, Germany, 2009.
- [5] L. K. Nielsen, *Rolling Stock Rescheduling in Passenger Railways: Applications in short-term planning and in disruption management*, Erasmus University Rotterdam, 2011.
- [6] B. Pender, G. Currie, A. Delbosc, and N. Shiwakoti, "Planning for the unplanned: An international review of current approaches to service disruption management of railways," in *Proceedings of the 35th Australasian Transport Research Forum, ATRF 2012*, aus, September 2012.
- [7] R. Silva, S. M. Kang, and E. M. Airoidi, "Predicting traffic volumes and estimating the effects of shocks in massive transportation systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 18, pp. 5643–5648, 2015.
- [8] Y. Chung, "A methodological approach for estimating temporal and spatial extent of delays caused by freeway accidents," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1454–1461, 2012.
- [9] C. Liu, S. Zhang, H. Wu, and Q. Fu, "A Dynamic Spatiotemporal Analysis Model for Traffic Incident Influence Prediction on Urban Road Networks," *ISPRS International Journal of Geo-Information*, vol. 6, no. 11, p. 362, 2017.
- [10] J. Wu, M. Liu, H. Sun, T. Li, Z. Gao, and D. Z. W. Wang, "Equity-based timetable synchronization optimization in urban subway network," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 1–18, 2015.
- [11] X. Guo, J. Wu, H. Sun, R. Liu, and Z. Gao, "Timetable coordination of first trains in urban railway network: a case study of Beijing," *Applied Mathematical Modelling: Simulation and Computation for Engineering and Environmental Systems*, vol. 40, no. 17–18, pp. 8048–8066, 2016.
- [12] X. Guo, H. Sun, J. Wu, J. Jin, J. Zhou, and Z. Gao, "Multiperiod-based timetable optimization for metro transit networks,"

Transportation Research Part B: Methodological, vol. 96, pp. 46–67, 2017.

- [13] X. Yang, X. Li, B. Ning, and T. Tang, “An optimisation method for train scheduling with minimum energy consumption and travel time in metro rail systems,” *Transportmetrica B: Transport Dynamics*, vol. 3, no. 2, pp. 79–98, 2015.
- [14] J. Yang, J. G. Jin, J. Wu, and X. Jiang, “Optimizing Passenger Flow Control and Bus-Bridging Service for Commuting Metro Lines,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 6, pp. 458–473, 2017.
- [15] X. Yang, A. Chen, B. Ning, and T. Tang, “Bi-objective programming approach for solving the metro timetable optimization problem with dwell time uncertainty,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 97, pp. 22–37, 2017.
- [16] X. Yang, A. Chen, J. Wu, Z. Gao, and T. Tang, “An energy-efficient rescheduling approach under delay perturbations for metro systems,” *Transportmetrica B: Transport Dynamics*, pp. 1–15, 2018.
- [17] H. Yin, J. Wu, H. Sun, L. Kang, and R. Liu, “Optimizing last trains timetable in the urban rail network: social welfare and synchronization,” *Transportmetrica B: Transport Dynamics*, pp. 1–25, 2018.
- [18] J. Yin, Y. Wang, T. Tang, J. Xun, and S. Su, “Metro train rescheduling by adding backup trains under disrupted scenarios,” *Frontiers of Engineering Management*, vol. 4, no. 4, p. 418, 2017.
- [19] 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.
- [20] Y. Wang, J. Guo, G. Currie, A. Ceder, W. Dong, and B. Pender, “Bus bridging disruption in rail services with frustrated and impatient passengers,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2014–2023, 2014.
- [21] J. G. Jin, K. M. Teo, and L. Sun, “Disruption response planning for an urban mass rapid transit network[C],” in *Transportation Research Board (TRB) 92nd Annual Meeting*, pp. 1–18, Transportation Research Board, Washington DC, USA, 2013.
- [22] J. G. Jin, K. M. Teo, and A. R. Odoni, “Optimizing bus bridging services in response to disruptions of urban transit rail networks,” *Transportation Science*, vol. 50, no. 3, pp. 790–804, 2016.
- [23] K. Lu, B. Han, F. Lu, and Z. Wang, “Urban Rail Transit in China: Progress Report and Analysis (2008–2015),” *Urban Rail Transit*, vol. 2, no. 3-4, pp. 93–105, 2016.

Research Article

A Crash Surrogate Metric considering Traffic Flow Dynamics in a Motorway Corridor

Xu Wang¹ and Kai Liu² 

¹Griffith School of Engineering, Griffith University, Gold Coast, QLD 4222, Australia

²School of Transportation and Logistics, Dalian University of Technology, Dalian 116024, China

Correspondence should be addressed to Kai Liu; liukai@dlut.edu.cn

Received 20 February 2018; Revised 22 April 2018; Accepted 6 May 2018; Published 27 June 2018

Academic Editor: Zhiyuan Liu

Copyright © 2018 Xu Wang and Kai 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.

We proposed a new crash surrogate metric, i.e., the maximum disturbance that a car following scenario can accommodate, to represent potential crash risks with a simple closed form. The metric is developed in consideration of traffic flow dynamics. Then, we compared its performance in predicting the rear-end crash risks for motorway on-ramps with other two surrogate measures (time to collision and aggregated crash index). To this end, a one-lane on-ramp of Pacific Motorway, Australia, was selected for this case study. Due to the lack of crash data on the study site, historical crash counts were merged according to levels of service (LOS) and then converted into crash rates. In this study, we used the societal risk index to represent the crash surrogate indicators and built relationships with crash rates. The final results show that (1) the proposed metric and aggregated crash index are superior to the time to collision in predicting the rear-end crash risks for on-ramps; (2) they have a relatively similar performance, but due to the simple calculation, the proposed metric is more applicable to some real-world cases compared with the aggregated crash index.

1. Introduction

According to the World Health Organization [1], approximately 1.24 million people died and over 50 million were injured in road crashes all over the world. More importantly, road crashes have been the leading cause of death for young people aged 15–29 years. In Australia, the social cost of road crashes has been estimated as a devastating AU\$ 27 billion per annum [2]. It has been well recognized that freeway crashes are much more dangerous than those on urban streets due to their severity [3–5]. According to United States Department of Transportation [6], the proportion of fatal crashes on freeways is significantly higher than that on other road types. Further, on-ramp areas of freeways may be prone to relatively more traffic crashes than other sections of freeways (Cheng et al., 2017) [7] because drivers have been facing a high competition for merging space in such areas [8, 9]. In this regard, capable solutions need to be sought in order to proactively assess or predict on-ramp crash.

Considerable research efforts have been carried out over the past fifty years on developing count-data regression models to predict crash frequency [10] and to investigate

crash casualty of different collision types [11, 12], most of which are purely dependent on statistical techniques. These count-data regression models, in accordance with distinct statistical assumptions, represent the relationship between number of crashes and its contributing factors which have little, if not none, consideration of traffic flow dynamics [13, 14]. Since the 1970s, some researchers began to use crash surrogate measure to evaluate road safety (e.g., [15–19]). Only recently has a consensus emerged concerning the definition of a crash surrogate, which is defined based on the relationship [20, 21]: the number of crashes expected to occur on an entity during a certain period of time (λ) = the number of crash surrogates occurring on an entity in that time (π) * crash-to-surrogate ratio for that entity (c), mathematically,

$$\lambda = \pi \cdot c. \quad (1)$$

A few crash surrogate metrics have been proposed and designed, including the time to collision (TTC), the deceleration rate to avoid crash (DRAC), and the crash potential index (CPI). However, as proposed by Kuang et al. [22], these models are incapable of representing crash surrogates on freeways, especially saturated freeways where

a minor disturbance can result in a rear-end crash due to very high speeds and small headways. For example, if the time headway between two consecutive vehicles on a freeway is 0.5 seconds and their speeds are equal at 110 km/hr, all these crash surrogate metrics will identify it as a safe scenario. In other words, the abovementioned crash surrogate metrics fail to assess crash risks of this particular car following scenario, which may lead to erroneous judgement. In this regard, Kuang et al. [22] proposed a tree-structured crash surrogate metric by imposing a hypothetical disturbance to the leading vehicle. An aggregated crash index (ACI) was proposed to combine eight possible scenarios caused by the imposed hypothetical disturbance. According to the validation, the ACI outperforms the traditional TTC based surrogate metrics in representing freeway rear-end crash risks.

The biggest disadvantage of this tree-structured crash surrogate measure is that a closed form is not available due to this rather complicated tree structure. It, unfortunately, limits the metric's applicability to deal with real-world problems. For example, the model is naturally applicable to optimize the traffic operations of connected and automated vehicles in order to achieve the highest safety level. Unfortunately, as there is no closed form for ACI, the traditional optimization models cannot be used and only simulation based optimization models can be considered as an alternative. As such, it is difficult to analyze the analytical properties of the optimization results.

In this paper, we proposed a new concept of traffic state vulnerability to develop a simplified crash surrogate metric (SCSM). Traffic state vulnerability is defined as the maximum disturbance that a car following scenario can accommodate. With this new concept, the SCSM with a closed form was proposed. We further compared the performances of this new surrogate metric, ACI, and the conventional TTC based surrogate metrics for on-ramps. According to a comparative analysis, although the SCSM has a much simpler form, it has more or less similar performance compared to ACI, which outperforms the TTC. The rest of the paper is organized as follows. Section 2 briefs the TTC and ACI. The SCSM is introduced in Section 3. A comparative analysis is carried out in Section 4. Section 5 concludes this study and points out some future research directions and possible application of the new metric.

2. Literature Review

Multiple researchers have made great efforts to proactively predict accident risks so that various surrogate measures of safety have been designed and developed over past decades. Among them, the most widely used surrogate measure of TTC is defined as the time remaining until a collision will occur between two vehicles if the collision course and speed difference are maintained [23], or mathematically,

$$\text{TTC} = \begin{cases} \frac{D_{l-f}}{v_f - v_l}, & \text{if } v_f > v_l \\ \infty, & \text{otherwise,} \end{cases} \quad (2)$$

where v_f is the speed of follower's vehicle in m/s, v_l is the speed of leader's vehicle in m/s, and D_{l-f} is the gap between the two vehicles in metre. All TTC values need to compare with its threshold which has ranged from 1.5 s to 4 s in previous studies [24]. Only those car following scenarios with TTC values less than its threshold could be judged as risky.

The defects of the TTC contributed to the advance of other safety surrogate measures [25–27]. However, there are some shortcomings present in these methods. Firstly, boundary condition is improperly obeyed. For example, scenarios in which the speed of leading vehicle is higher than that of following vehicle are identified as safe situations. However, this condition boundary is not applicable to freeways with saturated traffic flow, as vehicles travelling at high speeds have relatively small differences in speeds. Accordingly, some potential risks on freeways may not be accurately measured. Secondly, the motorists' reaction time, as an essential parameter in crash risk assessment, fails to be considered in these surrogate metrics. Thirdly, the setting of threshold, like the TTC, lacks a specific judgement criterion. To address the drawbacks, Kuang et al. [22] proposed a safety surrogate measure called ACI which can better assess conflicts occurring on saturated freeways. Built on a probabilistic causal model, they first developed a tree-structured crash model being able to impose hypothetical disturbance. Eight possible conflict types under the model were recognized through four stepped condition levels.

The advantages of the ACI lie in two aspects. First, it was proposed for the risk assessment of car following scenarios on saturated freeways. Accordingly, it is able to eliminate false negative errors (in the saturated traffic flow, the scenarios where the speed of following vehicle is less than that of leading vehicle, but the speed difference is slight, are mistakenly viewed as safe). Second, the ACI incorporates two essential parameters like driver's reaction time and maximum available deceleration rate (MADR) in crash mechanism. They are invariably considered in the form of distribution instead of a fixed value.

However, its complex probability calculation limits its applicability to assess potential risks in the real world. As a result, the Monte Carlo method (MCM) may be the only method to simplify the complicated calculation processes involved in the ACI.

3. A Simplified Crash Surrogate Metric

Based on the notion of traffic state vulnerability, a new simplified crash surrogate metric (SCSM), with a closed form, is proposed. It is defined as the maximum disturbance that a traffic state could accommodate, namely, the ultimate capacity that a car following scenario can avoid a collision. It can be categorised into two scenarios based on the occurrence probability of a crash.

Scenario I. A crash will not occur under the condition that

$$v_f \leq v_l - \delta. \quad (3)$$

That is,

$$\delta \leq v_l - v_f, \quad (4)$$

where δ is the maximum interference that a traffic state can withstand.

Scenario II. A crash is possible to occur under the condition that

$$v_f > v_l - \delta. \quad (5)$$

That is,

$$\delta > v_l - v_f \quad (6)$$

Apparently, as v_f is greater than $v_l - \delta$, a crash will occur if both vehicles maintain the same speed. In other words, the modified time to collision with respect to disturbance δ can be calculated as

$$\text{TTC}_\delta = \frac{d_{l-f}}{v_f - (v_l - \delta)}. \quad (7)$$

To further evaluate the crash risks of a car following state, we need to compare the predicted remaining time to crash and the threshold of time to collision. If (8) holds, the following vehicle is able to timely decelerate and a crash could be avoided. Otherwise, a crash will occur. In this research, according to the previous researchers' experience [24, 28, 29], 3 s as the critical value can be suitable to assess crash risks on freeways with saturated flow.

$$\frac{d_{l-f}}{v_f - (v_l - \delta)} \geq \tau. \quad (8)$$

d_{l-f} can be estimated by

$$d_{l-f} = v_f h - L_l, \quad (9)$$

where h is the time headway of this car following scenario and L_l is the length of the leading vehicle. By substituting (9) into (8), we have

$$\frac{v_f h - L_l}{v_f - (v_l - \delta)} \geq \tau. \quad (10)$$

As $v_f - (v_l - \delta) > 0$, (10) is equivalent to

$$(v_l - v_f) < \delta \leq (v_l - v_f) + \frac{v_f h - L_l}{\tau}. \quad (11)$$

In other words, as long as a disturbance satisfies (11), a crash is avoidable; otherwise, a crash will occur. By combining the conditions for Scenarios I and II, we can conclude that a crash will not occur if

$$0 \leq \delta \leq (v_l - v_f) + \frac{v_f h - L_l}{\tau}. \quad (12)$$

Namely, $(v_l - v_f) + (v_f h - L_l)/\tau$ is the maximum disturbance that a car following scenario is able to accommodate. The greater δ is, the larger disturbance a traffic state can accommodate; namely, a car following scenario has better capacity to resist on-ramp rear-end crash risks.

TABLE 1: 24-hour based crash counts on the study on-ramp from year 2005 to year 2013.

Time period	LOS	Crash counts
00:00–01:00	A & B	0
01:00–02:00	A & B	1
02:00–03:00	A & B	1
03:00–04:00	A & B	0
04:00–05:00	A & B	0
05:00–06:00	D	1
06:00–07:00	E	2
07:00–08:00	E	3
08:00–09:00	D	1
09:00–10:00	D	1
10:00–11:00	C	0
11:00–12:00	C	1
12:00–13:00	C	1
13:00–14:00	C	0
14:00–15:00	C	2
15:00–16:00	D	2
16:00–17:00	D	2
17:00–18:00	E	3
18:00–19:00	D	1
19:00–20:00	C	1
20:00–21:00	C	0
21:00–22:00	A & B	1
22:00–23:00	A & B	0
23:00–00:00	A & B	0

4. A Comparative Analysis

4.1. Crash Data Processing. In this study, crash data were provided by Department of Transport and Main Roads (DTMR) and compiled based on annual daily hourly crash counts. By means of accident coordinates, all rear-end crashes occurring on the study on-ramp from year 2005 to year 2013 were extracted and listed in Table 1. Due to the lack of crash data, we determined to merge current crash counts based on levels of service (LOS). For this purpose, each time period in the 24 hours was marked with LOS based on the hourly traffic flow rates provided by DTMR from 15 to 19 April 2013 (5 workdays). Four groups of LOS were considered in this study: A&B, C, D, and E. According to HCM 2010, the determination of LOS highly depends on traffic volumes and capacity. Accordingly, LOS acted as an important role in this research. Based on the processed data, crash rates under the corresponding LOS from year 2005 to year 2013 can be calculated and presented in Table 2.

4.2. Simulation Model Development and Validation. In this research, VISSIM was used to simulate and reproduce down-to-earth traffic scenarios of the research segment [30]. An on-ramp of northbound Pacific Motorway, Queensland, adjacent to the largest shopping centre of suburb Nerang, was chosen as the research site, as it involves slight traffic delay and relatively many weaving manoeuvres during rush hours and

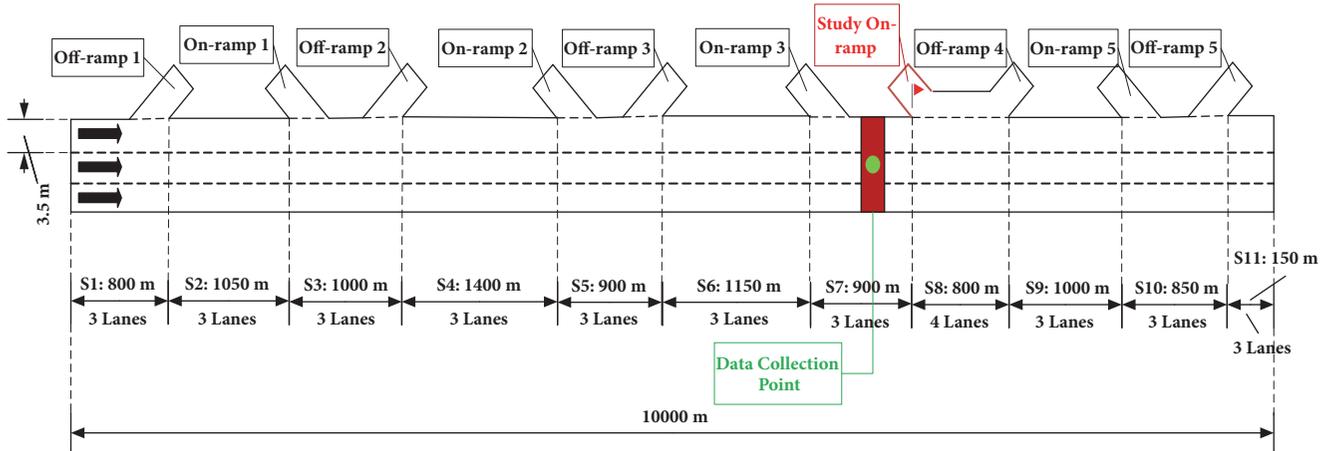


FIGURE 1: The sketch map of research area.

TABLE 2: Crash rate in the past nine years under the corresponding LOS.

LOS	The number of time periods	Crash counts	Crash rate (counts/hr)
A&B	8	3	0.38
C	7	5	0.71
D	6	8	1.33
E	3	8	2.67

has always been regarded as a bottleneck by Gold Coast City Council (GCCC). There were 36 data collection points set up on the 360-metre on-ramp with a spacing of 10 metres. To reproduce the impact of the up- and down-stream traffic flow on the research on-ramp, the simulation model was extended to a length of 10 kilometres. Additional 4 on-ramps and 5 off-ramps were involved within 10 kilometres. The sketch map of the research area is shown in Figure 1. There was no accident occurring during the data collection process and no major modification of the research on-ramp and its circumambient ramps from year 2005 to year 2013. All settings associated with geometrical characteristics need to be coded based on Google Earth Pro and behavioural parameters in the Wiedemann 99 model need to be calibrated with the real situations. According to the existing data, the average traffic flow rates under four different LOS can be derived through loop detectors setup at upstream 200 metres of off-ramp 1, which was used as traffic volume input in VISSIM. Similarly, the average number of merging traffic and diverging traffic for each on-ramp and off-ramp can also be obtained and shown in Table 3. Take traffic scenario 1 as an example. The traffic volume input for on-ramp 1, 2, 3, and 5 and the study on-ramp account for 7.1%, 6.5%, 4.6%, 4.1%, and 7.5% of their corresponding mainline traffic, respectively. Similarly, there are 5.4%, 5.3%, 7.2%, 6.3%, and 6.8% of mainline traffic diverging into off-ramp 1, 2, 3, 4, and 5.

To validate the effectiveness of VISSIM in generating dynamic vehicle behaviour, a bottleneck close to the study on-ramp in the motorway corridor was identified and used for

video recording. Four videos need to be recorded and to meet the following criterion: traffic volumes in the four videos are as close as possible to those generated by simulation models at the same location. Then, the traffic trajectories for 20 random consecutive vehicles in each video were extracted and compared with those generated from simulation models. Four error tests were herein carried out to evaluate the differences between the simulated results and the observed data: (1) root mean square error (RMSE), (2) root mean square percentage error (RMSPE), (3) mean percentage error (MPE), and (4) Theil's inequality coefficient (U), mathematically,

$$\begin{aligned}
 \text{RMSE} &= \sqrt{\frac{1}{N_0} \sum_{n=1}^{N_0} (y_n^s - y_n^0)^2} \\
 \text{RMSPE} &= \sqrt{\frac{1}{N_0} \sum_{n=1}^{N_0} \left(\frac{y_n^s - y_n^0}{y_n^0} \right)^2} \\
 \text{MPE} &= \frac{1}{N_0} \sum_{n=1}^{N_0} \left(\frac{y_n^s - y_n^0}{y_n^0} \right) \\
 U &= \frac{\sqrt{(1/N_0) \sum_{n=1}^{N_0} (y_n^s - y_n^0)^2}}{\sqrt{(1/N_0) \sum_{n=1}^{N_0} (y_n^s)^2} + \sqrt{(1/N_0) \sum_{n=1}^{N_0} (y_n^0)^2}},
 \end{aligned} \tag{13}$$

where y_n^s is the simulation value (speed) of the n th vehicle, y_n^0 is the field value (speed) of the n th vehicle, and N_0 is the number of vehicles observed or simulated. The error tests of speeds were shown in Table 4. For any error tests, the difference between simulated speeds and observed speeds is the greatest under LOS A&B, as the speeds for vehicles under LOS A&B are close to free speeds, which is difficult to capture through video recording. Besides, the selection of driving speeds is highly free under LOS A&B. As a result, a larger difference in speeds is inevitable. For a speed limit of 110 km/hr (30.56 m/s), an error of 3.37 m/s is in the acceptable range. The values for RMSPE are less than 15% across all groups. The largest value for MPE is positive 7.69%, which

TABLE 3: The traffic flow configuration for each on-ramp and off-ramp.

Traffic scenario	1	2	3	4
LOS	A&B	C	D	E
Freeway input (vehs/hr)	1830	3960	5220	6360
Off-ramp 1 output (vehs/r)	99 (5.4%)	245 (6.1%)	298 (5.7%)	426 (6.7%)
On-ramp 1 input (vehs/hr)	123 (7.1%)	238 (6.4%)	315 (6.4%)	362 (6.1%)
Off-ramp 2 output (vehs/hr)	98 (5.3%)	233 (5.9%)	346 (6.6%)	384 (6.1%)
On-ramp 2 input (vehs/hr)	114 (6.5%)	223 (6.0%)	303 (6.2%)	414 (7.0%)
Off-ramp 3 output (vehs/hr)	135 (7.2%)	334 (7.7%)	441 (8.5%)	512 (8.1%)
On-ramp 3 input (vehs/hr)	80 (4.6%)	189 (5.2%)	304 (6.4%)	401 (6.9%)
Study on-ramp input (vehs/hr)	136 (7.5%)	291 (7.6%)	405 (8.0%)	466 (7.5%)
Off-ramp 4 output (vehs/hr)	123 (6.3%)	276 (6.7%)	364 (5.2%)	488 (7.3%)
On-ramp 5 input (vehs/hr)	75 (4.1%)	165 (4.3%)	190 (3.9%)	279 (4.5%)
Off-ramp 5 output (vehs/hr)	130 (6.8%)	281 (7.0%)	329 (6.5%)	466 (7.2%)

TABLE 4: Error tests of speeds.

LOS group	RMSE (m/s)	RMSPE (%)	MPE (%)	U (%)
A&B	3.37	14.19	7.69	0.28
C	3.28	13.73	6.67	0.26
D	2.64	11.86	5.75	0.23
E	2.68	11.51	5.51	0.24

indicates that the simulated speeds are slightly overestimated by VISSIM compared with real speeds. Furthermore, U values are close enough to zero. The closer to zero the coefficient is, the smaller the difference in speeds is. Accordingly, we can conclude that VISSIM is able to well simulate a real traffic situation in terms of microscopic level.

4.3. Preliminary Test. To assess crash risks of all vehicles on the study on-ramp under four different LOS, the concepts of individual risk and societal risk were introduced. The individual risk (IR) is defined as the crash threat to an individual motorist, which is regarded as the likelihood of collision occurring to an individual traveller. The ACI, SCSM, and the difference between TTC and its threshold (if TTC is less than its threshold) all can be viewed as the IR. Building on the notion of IR, the societal risk (SR) index was proposed, which refers to the integrated risk of all individual risks to all of the affected drivers at on-ramp length L measured by surrogate metric j , mathematically denoted by

$$SR_j = \sum_{i=1}^M \int_0^L IR_{ij}(l) dl \approx \sum_{i=1}^M \sum_{l=0}^N IR_{ij}(l) \times i_{dc}, \quad (14)$$

where $IR_{ij}(l)$ represents the individual risk of the discrete scenario i at discrete length l measured by surrogate j , i_{dc} is the interval of two consecutive data collection points, and there are a total of N collection points in length L . In this study, 36 data collection points evenly spread on the 360-metre research on-ramp. Due to the complexity of the ACI, the Monte Carlo method (MCM) was used to simplify the calculation of crash probability. 10,000 random drivers'

reaction times and MADRs were generated by the MCM in each car following scenario. To make the numerical values more readable, we applied normalization method to process the results of the SCSM. Table 5 presents the SR represented by the TTC, ACI, and SCSM under four varying LOS. As crash data are limited, we cannot conclude which metric has a better ability to predict risks for the study on-ramp through a linear model. However, according to the proportional relation between the SR and crash rates, the performance of the SCSM is optimal, followed by the ACI and TTC. For the TTC, we generally compared the difference between the leading vehicle's speed and the following vehicle's speed to judge a car following scenario safe or risky. This judgement criterion is obviously unreasonable for assessing crash risks for on-ramps, since the speeds of leading vehicles in the merging traffic are larger than those of following vehicles in most cases. However, in a saturated merging traffic stream with smaller time headways and higher driving speeds, even though the abovementioned condition is met, a crash may occur. Accordingly, the traditional TTC is inappropriate for risk assessment for on-ramps. For the ACI, a robust probabilistic causal model enables it to capture any potential car following risks on the on-ramp. However, its complex calculation restricts its applicability in the real world. With the simplified calculation procedure and more or less similar performance to the ACI, the SCSM stand out of these three surrogate measures.

5. Conclusions and Limitations

Aiming at rear-end accidents occurring on on-ramps, this paper proposed a new concept of traffic state vulnerability in order to develop and assess the simplified crash surrogate metric (SCSM). As an upgraded version of the traditional time to collision (TTC), the SCSM not only features the same straightforward closed form as the traditional TTC, but also makes up for the shortcoming of the TTC that is unable to accurately assess crash risks in saturated traffic flow. In this study, we used it and another two surrogate measures, aggregated crash index (ACI), and TTC, to predict the crash risks for an on-ramp on Pacific Motorway in

TABLE 5: Surrogate-based societal risk under four LOS.

LOS	Crash counts	Crash rate (counts/hr)	SR represented by surrogate metrics		
			TTC	ACI	SCSM
A&B	3	0.38	0	119.07	29725
C	5	0.71	5.52	254.38	26540
D	8	1.33	15.91	333.67	24349
E	8	2.67	19.23	397.86	21937

Queensland, Australia. Based on a comparative analysis, the TTC based surrogate metric performed the worst, as it makes a concession to limited historical crash counts and boundary condition. Moreover, it is assessed based on the fact that the speeds for leading and following vehicles are invariable during the collision course, which is inconsistent with the real world. When the following vehicle takes necessary evasive actions to avoid a crash, the change in deceleration rates taken by the following vehicle is hard to capture. The SCSM tactfully takes advantage of the concept, the maximum disturbance that a traffic state could accommodate, thereby effectively evading the consideration for the complicated evasive actions involved in a car following scenario. In general, as a variation of the traditional TTC, the SCSM can better assess crash risks in the context of urban environment. The performance of the SCSM is more or less similar to that of the ACI. But considering the ability to resolve practical engineering issues, the SCSM is superior to the ACI. It can assist traffic agents in efficiently and precisely assessing rear-end crash risks for on-ramps.

The limitations of this research are summarized below. There is no validation for various crash surrogate metrics. Due to the restriction of historical crash counts on the research on-ramp, it is not convincing to only depend on a simple proportional relationship between the SR and crash rates to judge that the SCSM is superior to others. Therefore, we believe that the following solutions can to some extent make up for the limitation. Firstly, the similar on-ramps identified through VISSIM calibration and validation should be grouped together as the study on-ramps. Secondly, crash counts in each hour occurring on the subject on-ramps need to be extracted from historical crash record. Thirdly, based on the hourly simulation animation for these on-ramps, the results for various surrogate metrics can be obtained. Fourthly, the linear relationships between the SR and crash counts for each surrogate metric can be developed based on 24 points, which will be a sound validation to demonstrate that a certain crash surrogate measure is superior to others. Furthermore, the case study was only focused on an on-ramp on Pacific Motorway. In further study, we will apply these three surrogate measures to assess crash risks for other on-ramps on Pacific Motorway.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest for publication of this work.

Acknowledgments

The authors would like to acknowledge the funding provided by the National Natural Science Foundation of China (Grant no. 51378091), the National Natural Science Foundation of Liaoning Province, China (Grant no. 20170540187), and the Fundamental Research Funds for the Central Universities (Grant no. DUT18GJ204).

References

- [1] World Health Organization (WHO), *Global Report on Road Safety 2013: Supporting a Decade of Action*, WHO Press, World Health Organization, Geneva, Switzerland, 2013.
- [2] Department of Infrastructure and Regional Development (DIRD), "Road safety," 2013, <http://www.infrastructure.gov.au/roads/safety>.
- [3] Z. Liu, Q. Meng, and S. Wang, "Speed-based toll design for cordon-based congestion pricing scheme," *Transportation Research Part C: Emerging Technologies*, vol. 31, pp. 83–98, 2013.
- [4] Z. Liu, S. Wang, and Q. Meng, "Optimal joint distance and time toll for cordon-based congestion pricing," *Transportation Research Part B: Methodological*, vol. 69, pp. 81–97, 2014.
- [5] M. Zhou, X. Qu, and X. Li, "A recurrent neural network based microscopic car following model to predict traffic oscillation," *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 245–264, 2017.
- [6] United States Department of Transportation (US DOT), *Traffic Safety Fact 2005*, NHTSA's national center for statistic and analysis, 2005, <http://www-nrd.nhtsa.dot.gov/Pubs/810631.pdf>.
- [7] L. Zhang, C. Chen, J. Zhang, S. Fang, J. You, and J. Guo, "Modeling Lane-Changing Behavior in Freeway Off-Ramp Areas from the Shanghai Naturalistic Driving Study," *Journal of Advanced Transportation*, vol. 2018, pp. 1–10, 2018.
- [8] X. Qu, S. Wang, and J. Zhang, "On the fundamental diagram for freeway traffic: a novel calibration approach for single-regime models," *Transportation Research Part B: Methodological*, vol. 73, pp. 91–102, 2015.
- [9] X. Qu, J. Zhang, and S. Wang, "On the stochastic fundamental diagram for freeway traffic: Model development, analytical properties, validation, and extensive applications," *Transportation Research Part B: Methodological*, vol. 104, pp. 256–271, 2017.
- [10] H. Wu, L. Gao, and Z. Zhang, "Analysis of crash data using quantile regression for counts," *Journal of Transportation Engineering*, vol. 140, no. 4, Article ID 04013025, 2014.

- [11] C. Xu, H. Li, J. Zhao, J. Chen, and W. Wang, "Investigating the relationship between jobs-housing balance and traffic safety," *Accident Analysis & Prevention*, vol. 107, pp. 126–136, 2017.
- [12] C. Xu, Y. Wang, P. Liu, W. Wang, and J. Bao, "Quantitative risk assessment of freeway crash casualty using high-resolution traffic data," *Reliability Engineering & System Safety*, vol. 169, pp. 299–311, 2018.
- [13] D. Lord, S. D. Guikema, and S. R. Geedipally, "Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes," *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1123–1134, 2008.
- [14] S. Jin, X. Qu, C. Xu, D. Ma, and D. Wang, "An improved multi-value cellular automata model for heterogeneous bicycle traffic flow," *Physics Letters A*, vol. 379, no. 39, Article ID 23345, pp. 2409–2416, 2015.
- [15] H.-C. Chin and S.-T. Quek, "Measurement of traffic conflicts," *Safety Science*, vol. 26, no. 3, pp. 169–185, 1997.
- [16] D. Gettman and L. Head, "Surrogate safety measures from traffic simulation models," *Transportation Research Record*, no. 1840, pp. 104–115, 2003.
- [17] A. P. Tarko, "Use of crash surrogates and exceedance statistics to estimate road safety," *Accident Analysis & Prevention*, vol. 45, pp. 230–240, 2012.
- [18] C. Chai, X. Shi, and Y. D. Wong, "Estimating Safety Effects of Green-Man Countdown Devices at Signalized Pedestrian Crosswalk Based on Cellular Automata," *Journal of Advanced Transportation*, vol. 2017, pp. 1–13, 2017.
- [19] T. D. Chu, T. Miwa, and T. Morikawa, "Discrete choice models for gap acceptance at urban expressway merge sections considering safety, road geometry, and traffic conditions," *Journal of Transportation Engineering*, vol. 143, no. 7, Article ID 04017025, 2017.
- [20] E. Hauer, "Traffic conflicts and exposure," *Accident Analysis & Prevention*, vol. 14, no. 5, pp. 359–364, 1982.
- [21] A. Tarko, G. Davis, N. Saunier, T. Sayed, and S. Washington, *White paper: Surrogate Measures of Safety*, Committee on Safety Data Evaluation and Analysis (ANB20), 2009.
- [22] Y. Kuang, X. Qu, and S. Wang, "A tree-structured crash surrogate measure for freeways," *Accident Analysis & Prevention*, vol. 77, pp. 137–148, 2015.
- [23] J. Hayward, "Near miss determination through use of a scale of danger: Transportation Research Record," *Journal of the Transportation Research Board*, vol. 384, pp. 24–34, 1972.
- [24] R. Van der Horst, *Time-to-collision as a cue for decision-making in braking*, North-Holland, Amsterdam, 1991.
- [25] F. F. Saccomanno, F. Cunto, G. Guido, and A. Vitale, "Comparing safety at signalized intersections and roundabouts using simulated rear-end conflicts," *Transportation Research Record*, vol. 2078, pp. 90–95, 2008.
- [26] Q. Meng and X. Qu, "Estimation of vehicle crash frequencies in road tunnels," *Accident Analysis & Prevention*, vol. 48, pp. 254–263, 2012.
- [27] J. Weng, J.-Z. Zhu, X. Yan, and Z. Liu, "Investigation of work zone crash casualty patterns using association rules," *Accident Analysis & Prevention*, vol. 92, pp. 43–52, 2016.
- [28] X. Qu, Y. Kuang, E. Oh, and S. Jin, "Safety evaluation for expressways: a comparative study for macroscopic and microscopic indicators," *Traffic Injury Prevention*, vol. 15, no. 1, pp. 89–93, 2013.
- [29] X. Qu, Y. Yang, Z. Liu, S. Jin, and J. Weng, "Potential crash risks of expressway on-ramps and off-ramps: a case study in Beijing, China," *Safety Science*, vol. 70, pp. 58–62, 2014.
- [30] R. Fan, H. Yu, P. Liu, and W. Wang, "Using VISSIM simulation model and surrogate safety assessment model for estimating field measured traffic conflicts at freeway merge areas," *IET Intelligent Transport Systems*, vol. 7, no. 1, pp. 68–77, 2013.

Research Article

Optimal Design of Transportation Networks with Automated Vehicle Links and Congestion Pricing

Yipeng Ye ¹ and Hua Wang ^{1,2}

¹*School of Economics and Management, Tongji University, Shanghai 200092, China*

²*Department of Civil & Environmental Engineering, National University of Singapore, Singapore 117576*

Correspondence should be addressed to Hua Wang; hwang191901@gmail.com

Received 28 February 2018; Revised 5 May 2018; Accepted 17 May 2018; Published 27 June 2018

Academic Editor: David Z. W. Wang

Copyright © 2018 Yipeng Ye and Hua Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a bi-level network design model comprising automated vehicle (AV) links and congestion pricing to improve traffic congestion. As upper-level road planners strive to minimize total travel-time costs by optimizing both the network design and the congestion pricing, lower-level travelers make choices about their routes to minimize their individual travel costs. Our proposed model integrates a network design and congestion pricing to improve traffic congestion and we use a relaxation-based method to solve the model. We conducted a series of numerical tests to analyze the proposed model and solution method. Our results indicate that network design is more effective than congestion pricing when the AV market penetration is high and the opposite is true when AV penetration is low. More importantly, we find that a network design of automated vehicle links with congestion pricing is superior to a single network design or congestion pricing, especially when both AVs and conventional vehicles have a relatively large market penetration.

1. Introduction

Urban transportation plays an important role in economic activity throughout the world. However, road congestion has become a serious problem in many metropolitan cities and has led to various problems, including economic inefficiency, traffic accidents, pollution, and excessive energy consumption. Network design and congestion pricing are classic and effective instruments for solving traffic congestion and have been studied for decades. Some studies have proposed network design strategies [1, 2]. Also, some comprehensive reviews have summarized the results of studies of the transportation network design problem (NDP) [3, 4]. Other studies worldwide have proposed and tested practical congestion pricing [5–7].

Automated vehicles (AV) have attracted increasing attention in recent years and they are expected to improve the future safety levels and traffic congestion of existing urban transportation systems through improvements in traffic safety, road capacity, travel time, energy consumption, and

pollution levels. Two of the anticipated advantages (relative to human-driver vehicle control) are increased road network capacity and the freeing up of driver-occupant time to engage in leisure activities or economically productive (nondriving) tasks [8]. Jiang [9] studied optimal signal design for mixed equilibrium networks with autonomous and regular vehicles. With the rapid increase of car ownership, the problem of traffic emission and lack of land for parking becomes a serious issue [10], and AV can substantially reduce traffic emission and save land for parking. AVs can also help traffic safety. Zhu and Ukkusuri [11] indicated that connected vehicle technology would potentially reduce 81% of all-vehicle target crashes, 83% of all light-vehicle target crashes, and 72% of all heavy-truck target crashes annually. Tientrakool et al. [12] and Chen et al. [13] stated that the capacity of AV links may be approximately triple due to the benefits resulting from vehicle-to-vehicle communication. It follows that congestion pricing may become unnecessary when conventional vehicles (CV) are completely phased out. Despite these exciting developments regarding AVs, CVs cannot be phased out

immediately and will continue to be widely used for many years. As such, it is necessary to consider the NDP and congestion pricing during this transitional period characterized by heterogeneous AV and CV travel modes.

Among the transportation NDP and congestion pricing studies addressing AV and CV modes, few focus on the simultaneous optimization of the transportation system NDP and congestion pricing. An NDP with congestion pricing can be viewed as a Stackelberg game between the road planner and travelers, in which the road planner acts as the leader and travelers are the followers. In this paper, we investigate transportation networks involving both AV and CV modes and present an optimal and simultaneous network design and congestion pricing for these networks. We assume that some links in the network belong to the candidate link set and all links in this set can be converted to AV links or remain as regular links. AV links would allow only AVs to operate on them and their capacities could triple, as reported in Tientrakool et al. [12] and Chen et al. [13]. Regular links in the candidate link set allow both AVs and CVs to operate on them. To induce traffic flow and promote AV market penetration, the road planner can implement congestion pricing for CVs on the regular links in the candidate link set. Once the AV links are established and congestion pricing is deployed, we assume that AVs and CVs will follow the Wardrop equilibrium principle with respect to route choice to minimize their individual travel costs and yield a multiclass user equilibrium (UE). The results of this study show that the proposed NDP with congestion pricing can be an effective method for minimizing total travel time cost in a network comprising both AV and CV modes. More importantly, we find that NDP with congestion pricing outperforms a single NDP or congestion pricing, especially when both AVs and CVs have fairly large market penetrations. This means that the NDP with congestion pricing is an effective approach for alleviating traffic congestion in the transitional period characterized by heterogeneous AV and CV travel modes.

1.1. Literature Review. Although AVs are expected to greatly improve the safety and efficiency of transportation networks, proper management instruments must be applied in the transitional period during which there are heterogeneous AV and CV travel modes. Both the NDP and the congestion pricing are considered to be effective instruments for alleviating traffic congestion and they are typically employed to maximize network performance (e.g., minimize total travel-time cost, maximize social welfare) while travelers' route choices are taken into account. The literature includes several comprehensive reviews of NDP, including those by Yang and H. Bell [4] and Farahani et al. [3]. Congestion pricing has also been extensively studied by both road managers and researchers and has long been considered to be an effective method for managing traffic demand and increasing tax revenue. Zhang and Yang [14] developed a method of simultaneously determining optimal locations and toll levels of cordon-based congestion pricing. Zhang et al. [15] studied the cooperation and competition congestion pricing policy among multiple regions. We refer interested readers to Yang and Huang [16] for details regarding developments in road

pricing research. In this section, we briefly summarize recent studies related to our work, including AV research, joint road transportation management, and traffic optimization methods.

(1) *Research on AVs.* Many studies have highlighted the potential social benefits of AVs. Le Vine et al. [8] identified two benefits of road vehicle automation: (I) increased road capacity and (II) the freeing up the drivers' in-car time for a wide range of economically valuable activities. Chen et al. [13] indicated that there are potential benefits of AVs to increase traffic safety, driver productivity, road capacity, and travel speed. Ji et al. [17] pointed out that connected AVs can be easily operated due to the application of information technology in AVs.

Although existing AV developments and emerging innovations in AV technology indicate the huge potential for improving traffic safety and congestion, applications of AV technologies are as yet limited. In general, AV technologies are being developed to sense and make assessments about the environment in which the AV is traveling (e.g., other travelers, road signs, and traffic flow density) and to initiate the appropriate actions in response. However, these assessments fully depend on the proper functioning of the sensory devices, which remain in the research-and-development phase with numerous ongoing experiments [18]. Some studies have sought to improve AV performance via technical innovation. Häne et al. [19] developed an obstacle detection system for self-driving cars, using monocular cameras and wheel odometry. Aparicio et al. [20] aimed to improve the effectiveness of safety systems for active vulnerable road users (VRUs) that are currently on the market by expanding the scope of scenarios addressed by these systems and improving the overall system performance.

Although currently immature AV technologies are being rapidly developed, it will be many years before AVs are widely adopted. As such, a long-term period of mixed AV and CV traffic flow is inevitable. Some researchers have suggested the expansion of dedicated AV areas to improve overall system traffic performance. For example, several studies have suggested the conversion of some traffic lanes into dedicated AV lanes to alleviate traffic congestion and improve traveler safety, since existing managed lanes are already equipped with communication systems [21]. Chen et al. [13] proposed the application of a time-dependent AV-lane design model on a general network for both AVs and CVs. Godsmark and Kakkar [22] proposed that AV areas can be designed to maximize the benefits realized from AVs, as well as to promote AV adoption. In addition, Chen et al. [23] proposed a mathematical framework for designing an optimal AV zone to achieve these goals.

(2) *Joint Managements for Road Transportation.* There has been an abundance of literatures published on the NDP and congestion pricing over the past half century. Readers can refer to Yang and H. Bell [4], Farahani et al. [3], and Yang and Huang [16] for comprehensive reviews and detailed discussions regarding the NDP and congestion pricing. Researchers also studied other kinds of joint implementation, such as

joint management of road tolls and road credits. Jiang et al. [24] investigated simultaneous optimization of road tolls and tradable road credits in public-private mixed traffic networks. Wang and Zhang [7] examined the joint implementation of tradable credit and road pricing in public-private partnership networks considering UE-CN mixed equilibrium behaviors. In this section, we review recent studies related to the joint management of road transportation.

Considering the development of network modelling to simultaneously determine network design strategy and road pricing, Koh et al. [25] studied joint optimal pricing and road capacity investment problems ranging from policy to technology perspectives. Zhang and van Wee [26] proposed a simultaneous toll-location and toll-level optimization model to maximize the network reserve capacity, which differs substantially from previous studies that aimed to minimize total travel time or maximize total social welfare. Wang et al. [27] proposed a bi-level programming model comprising a joint optimal-link-based tradable-credit-charging scheme and road capacity improvement. Their study results indicated that the synergistic effect of link-based credit charging and road capacity improvement enhanced urban transportation network performance. Xu et al. [28] proposed a discrete network design model with a road pricing scheme for minimizing total travel time under budgetary constraints. This optimal road capacity improvement and toll-level scheme considered the available budget levels and toll revenues.

(3) *Traffic Optimization Methods.* In recent years, a number of studies have focused on the development of algorithms for NDP and road pricing, including sensitivity-analysis-based [9, 29], cutting constraint [30], linearization [31, 32], and relaxation [27, 33] algorithms. Some metaheuristic algorithms have also been applied to solve these kinds of problems. For example, Meng et al. [5] proposed the hybrid genetic algorithm- (GA-) cost averaging (CA) method to solve the optimal-distance-based toll design problem for cordon-based congestion pricing schemes. Sun et al. [34] used a particle swarm optimization algorithm to design reliability-based traffic networks with advanced traveler information systems.

1.2. *Positioning and Objective.* Some researchers have considered the use of infrastructure adaption planning for AVs to reduce traffic congestion and improve traveler safety [21, 22, 35]. However, these works offered no integrated methods for solving traffic congestion. With this paper, we fill this gap by proposing a simultaneous optimization model of the NDP and congestion pricing for transportation networks with mixed AV and CV flow.

Although network design and congestion pricing have long been subjects of study, to the best of our knowledge, no studies have offered a specialized solution method for a discrete NDP with continuous pricing. In this paper, we propose a relaxation-based method for solving this kind of problem.

Specifically, our objectives in this paper are twofold. First, we propose a bi-level network design model of AV links and congestion pricing for improving traffic congestion, while

considering traveler behavior with respect to route choice. Secondly, we propose a relaxation-based method for solving the above model, which can also be applied to solve a discrete NDP with continuous pricing.

1.3. *Challenges, Our Solution Method, and Contributions.* To achieve the above objectives, we have formulated an NDP with congestion pricing using bi-level programming, while also considering traveler behavior regarding route choice. Not surprisingly, this model is computationally challenging. The first difficulty is that the bi-level model is a NP-hard problem, even if it involves bi-level linear programming [36]. The second difficulty is that the variables in the upper level of the bi-level programming contain integer (network design variables) and continuous (pricing variables) aspects, which some previously developed bi-level programming algorithms for the NDP and pricing (e.g., Wang et al. [33], Wang et al. [27]) are unable to solve.

To tackle these computational challenges, we decomposed the bi-level programming process into a number of subproblems. The subproblems for determining the upper bounds can be solved by the relaxation algorithm proposed by Wang et al. [27]. The subproblems for determining the lower bounds can be solved by the outer-approximation algorithm proposed by Wang et al. [33]. As the number of iterations increases, the gap between the lower and upper bounds decreases sharply and finally converges to zero.

Contributions. This paper makes the following contributions:

- (1) We propose a bi-level programming to formulate the NDP with congestion pricing for transportation networks with mixed AV and CV traffic flow. To the best of our knowledge, this is the first model that incorporates both network design and congestion pricing for this kind of network.
- (2) We propose a relaxation-based method for solving the model, which can also be applied to solve a discrete NDP with continuous pricing.

Managerial Insights. This paper also contributes to the following managerial insights:

- (1) If we compare congestion pricing with network design, we find that congestion pricing is effective when AV market penetration is low and network design is effective when AV penetration is high. As such, road planners must apply different strategies to network with different AV market penetrations.
- (2) Network design becomes more effective as the origin-destination (OD) demand is increasing, so it follows that AV promotion and AV link design may be the solution for alleviating traffic congestion when traffic demand continually increases.
- (3) In the transitional period characterized by heterogeneous AV and CV travel modes, the NDP with congestion pricing strongly outperforms either network design or congestion pricing when both AVs and CVs have fairly large market penetrations. This

finding indicates that the NDP with congestion pricing represents an effective management method for the transitional period when both AVs and CVs are operating.

This paper is organized as follows. In Section 2, we list the symbols used and propose our bi-level model for transportation systems in which both AV and CV modes are being used. In Section 3, we propose the use of the relaxation-based method for solving the bi-level model. In Section 4, we describe the numerical tests we conducted to analyze the proposed model and solution method. In Section 5, we conclude this paper with a brief summary and suggestions for future work.

2. Model Formulation

The network design problem (NDP) can be represented as a Stackelberg game between the road planner and travelers. The road planner can induce but not control traveler route choices. This game can be formulated as a bi-level optimization model, in which the planner in the upper level determines the network design and congestion pricing to minimize the total travel cost in the transportation network and the travelers determine their routes to minimize their individual travel costs. We list the set, parameter, and decision variable symbols as follows:

Sets

- N : Set of nodes in the transportation network
- A : Set of links in the transportation network
- \tilde{A} : Set of candidate AV links in the transportation network, $\tilde{A} \subseteq A$
- W : Set of origin–destination (OD) pairs in the transportation network
- K_w^{AV} : Path set for AV mode between OD pair $w \in W$
- K_w^{CV} : Path set for CV mode between OD pair $w \in W$

Parameters

- $\delta_{a,k}$: A binary coefficient which equals 1 if path k uses link $a \in A$; otherwise $\delta_{a,k} = 0$
- c_a : Link capacity of link $a \in A \setminus \tilde{A}$
- c_a^1 : Link capacity of link $a \in \tilde{A}$ if link a is set to be AV link
- c_a^2 : Link capacity of link $a \in \tilde{A}$ if link a remains to be regular link
- D_w^{AV} : Demands of AV mode for OD pair $w \in W$
- D_w^{CV} : Demands of CV mode for OD pair $w \in W$
- $t_a(v_a, C_a)$: Travel-time cost function on link $a \in A$ when the link flow and capacity are v_a and C_a , respectively
- L : A positive large number

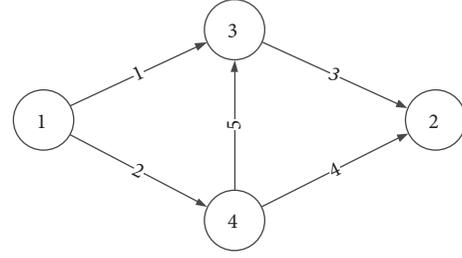


FIGURE 1: Schematic of a network with four nodes and five links.

Decision Variables

- x_a : A binary decision variable. $x_a = 1$ if link $a \in \tilde{A}$ is set to be AV link. Otherwise $x_a = 0$
- τ_a : Link-based pricing on $a \in \tilde{A}$ for CV
- \mathbf{x} : A vector defined as $\mathbf{x} = (x_a, a \in \tilde{A})$
- $\boldsymbol{\tau}$: A vector defined as $\boldsymbol{\tau} = (\tau_a, a \in \tilde{A})$
- v_a^{AV} : Link flow of AV mode on link a
- v_a^{CV} : Link flow of CV mode on link a
- v_a : Link flow on link a , which equals to $v_a^{AV} + v_a^{CV}$
- \mathbf{v} : A vector defined as $\mathbf{v} = (v_a, a \in A)$
- f_k^{AV} : Path flow of AV mode on path k
- f_k^{CV} : Path flow of CV mode on path k .

In this study, we assume that the road planner will aim at minimizing the total travel-time cost in the network by the network design and congestion pricing. After being presented with a designed network and congestion pricing for CVs, all travelers will strive to minimize their individual travel costs and their route choices can be characterized by the UE principle.

In addition, we assume that all links in the network are regular. However, links in candidate link set \tilde{A} can be converted to AV links and only AVs can operate on AV links. We note that the capacity of a link can be tripled by converting it from a regular to an AV link [12, 13]. It follows that pure AV flow will involve much less travel time when it passes through AV links. If some links in the candidate link set \tilde{A} are not converted to AV links, then the road planner can implement congestion pricing for CVs on these links. For example, we use the schematic in Figure 1 to illustrate network design and congestion pricing for a network comprising four nodes and five links.

In this network, we consider link 1 and link 4 to be candidate links that can be converted to AV links. That is, $A = \{1, 2, 3, 4, 5\}$ and $\tilde{A} = \{1, 4\}$. However, we only choose link 1 to be an AV link and $x_1 = 1$ and $x_4 = 0$. It follows that link 4 is still a regular link and can be a tolled link for CVs that run on it.

Based on the above assumptions, we can formulate the NDP bi-level programming (BLP) as shown below:

[BLP]

$$\begin{aligned} \min_{\mathbf{x}, \boldsymbol{\tau}} & \sum_{a \in A \setminus \bar{A}} t_a(v_a(\mathbf{x}, \boldsymbol{\tau}), c_a) v_a(\mathbf{x}, \boldsymbol{\tau}) \\ & + \sum_{a \in \bar{A}} x_a t_a(v_a(\mathbf{x}, \boldsymbol{\tau}), c_a^1) v_a(\mathbf{x}, \boldsymbol{\tau}) \\ & + \sum_{a \in \bar{A}} (1 - x_a) t_a(v_a(\mathbf{x}, \boldsymbol{\tau}), c_a^2) v_a(\mathbf{x}, \boldsymbol{\tau}) \end{aligned} \quad (1)$$

s.t.

$$x_a \in \{0, 1\}, \quad \forall a \in \bar{A}, \quad (2)$$

$$0 \leq \tau_a \leq L(1 - x_a), \quad \forall a \in \bar{A}, \quad (3)$$

where flow \mathbf{v} solves the following UE problem using link tolls and flow constraints for CVs:

$$\begin{aligned} \min_{\mathbf{v}} & \sum_{a \in A \setminus \bar{A}} \int_0^{v_a} t_a(\theta, c_a) d\theta + \sum_{a \in \bar{A}} x_a \int_0^{v_a} t_a(\theta, c_a^1) d\theta \\ & + \sum_{a \in \bar{A}} (1 - x_a) \int_0^{v_a} t_a(\theta, c_a^2) d\theta + \sum_{a \in \bar{A}} \tau_a v_a^{\text{CV}} \end{aligned} \quad (4)$$

s.t.

$$\sum_{k \in K_w^{\text{AV}}} f_k^{\text{AV}} = D_w^{\text{AV}}, \quad \forall w \in W, \quad (5)$$

$$\sum_{k \in K_w^{\text{CV}}} f_k^{\text{CV}} = D_w^{\text{CV}}, \quad \forall w \in W, \quad (6)$$

$$v_a^{\text{AV}} = \sum_{w \in W} \sum_{k \in K_w^{\text{AV}}} \delta_{a,k} f_k^{\text{AV}}, \quad \forall a \in A, \quad (7)$$

$$v_a^{\text{CV}} = \sum_{w \in W} \sum_{k \in K_w^{\text{CV}}} \delta_{a,k} f_k^{\text{CV}}, \quad \forall a \in A, \quad (8)$$

$$v_a = v_a^{\text{AV}} + v_a^{\text{CV}}, \quad \forall a \in A, \quad (9)$$

$$f_k^{\text{AV}} \geq 0, \quad \forall k \in K_w^{\text{AV}}, \quad \forall w \in W, \quad (10)$$

$$f_k^{\text{CV}} \geq 0, \quad \forall k \in K_w^{\text{CV}}, \quad \forall w \in W. \quad (11)$$

$$x_a v_a^{\text{CV}} = 0, \quad \forall a \in \bar{A}. \quad (12)$$

In the above BLP, (1) and (4) are objective functions of upper-level and lower-level programs, respectively. The road planner in the upper level determines the network design variable \mathbf{x} and congestion pricing variable $\boldsymbol{\tau}$, based on the responses of travelers being stable at UE. Constraint (2) ensures that any candidate AV link can be either an AV or regular link. Constraint (3) ensures that only regular links can be tolled links for CVs. Constraints (5)-(11) ensure flow balance in the network. Constraint (12) ensures that only regular links allow CVs to run on them.

Suppose that the link travel-time cost function is the following Bureau of Public Roads (BPR) function $t_a(v_a, C_a) = t_a^0(1 + 0.15(v_a/C_a)^4)$, where t_a^0 is the free-flow travel-time

cost of link a . For a given network design \mathbf{x} and congestion pricing $\boldsymbol{\tau}$, the lower-level programming is strictly convex and has a unique optimal solution for link flows. It follows that the lower-level programming can be replaced with its Karush–Kuhn–Tucker (KKT) conditions. The KKT conditions for the lower-level programming are as follows:

$$\sum_{k \in K_w^{\text{AV}}} f_k^{\text{AV}} = D_w^{\text{AV}}, \quad \forall w \in W, \quad (13)$$

$$\sum_{k \in K_w^{\text{CV}}} f_k^{\text{CV}} = D_w^{\text{CV}}, \quad \forall w \in W, \quad (14)$$

$$v_a^{\text{AV}} = \sum_{w \in W} \sum_{k \in K_w^{\text{AV}}} \delta_{a,k} f_k^{\text{AV}}, \quad \forall a \in A, \quad (15)$$

$$v_a^{\text{CV}} = \sum_{w \in W} \sum_{k \in K_w^{\text{CV}}} \delta_{a,k} f_k^{\text{CV}}, \quad \forall a \in A, \quad (16)$$

$$v_a = v_a^{\text{AV}} + v_a^{\text{CV}}, \quad \forall a \in A, \quad (17)$$

$$x_a v_a^{\text{CV}} = 0, \quad \forall a \in \bar{A}, \quad (18)$$

$$\begin{aligned} & \sum_{a \in \bar{A}} x_a t_a(v_a, c_a^1) \delta_{a,k} + \sum_{a \in \bar{A}} (1 - x_a) t_a(v_a, c_a^2) \delta_{a,k} \\ & + \sum_{a \in A \setminus \bar{A}} t_a(v_a, c_a) \delta_{a,k} - c_w^{\text{AV}} \geq 0, \end{aligned} \quad (19)$$

$$f_k^{\text{AV}} \geq 0, \quad \forall k \in K_w^{\text{AV}}, \quad \forall w \in W,$$

$$\begin{aligned} & \left(\sum_{a \in \bar{A}} x_a t_a(v_a, c_a^1) \delta_{a,k} + \sum_{a \in \bar{A}} (1 - x_a) t_a(v_a, c_a^2) \delta_{a,k} \right. \\ & \left. + \sum_{a \in A \setminus \bar{A}} t_a(v_a, c_a) \delta_{a,k} - c_w^{\text{AV}} \right) f_k^{\text{AV}} = 0, \end{aligned} \quad (20)$$

$$\forall k \in K_w^{\text{AV}}, \quad \forall w \in W,$$

$$\begin{aligned} & \sum_{a \in \bar{A}} x_a t_a(v_a, c_a^1) \delta_{a,k} + \sum_{a \in \bar{A}} (1 - x_a) t_a(v_a, c_a^2) \delta_{a,k} \\ & + \sum_{a \in \bar{A}} (\tau_a + \lambda_a x_a) \delta_{a,k} + \sum_{a \in A \setminus \bar{A}} t_a(v_a, c_a) \delta_{a,k} - c_w^{\text{CV}} \end{aligned} \quad (21)$$

$$\geq 0, \quad f_k^{\text{CV}} \geq 0, \quad \forall k \in K_w^{\text{CV}}, \quad \forall w \in W,$$

$$\begin{aligned} & \left(\sum_{a \in \bar{A}} x_a t_a(v_a, c_a^1) \delta_{a,k} + \sum_{a \in \bar{A}} (1 - x_a) t_a(v_a, c_a^2) \delta_{a,k} \right. \\ & \left. + \sum_{a \in \bar{A}} (\tau_a + \lambda_a x_a) \delta_{a,k} + \sum_{a \in A \setminus \bar{A}} t_a(v_a, c_a) \delta_{a,k} \right. \end{aligned} \quad (22)$$

$$\left. - c_w^{\text{CV}} \right) f_k^{\text{CV}} = 0, \quad \forall k \in K_w^{\text{CV}}, \quad \forall w \in W.$$

3. Relaxation-Based Method for Network Design with Congestion Pricing

Solving the NDP with congestion pricing is a challenge since the decision variables of the road planner contain both discrete (AV link design) and continuous (pricing) variables. Wang et al. [33] proposed bi-level programming and a relaxation method for the discrete network design problem (DNBP). Our work extends the DNBP approach since we also consider continuous pricing variables. Here, we combine efficient and effective relaxation algorithms for the NDPs addressed in Wang et al. [27] and Wang et al. [33] and propose a relaxation-based method to solve the network design with congestion pricing.

Let

$$\begin{aligned}\Omega_{\mathbf{x},\boldsymbol{\tau}} &= \{\mathbf{x}, \boldsymbol{\tau} : \mathbf{x} \text{ and } \boldsymbol{\tau} \text{ satisfy constraints (2) – (3)}\}, \\ \Omega_{\mathbf{v}}^{\text{RP}} &= \{\mathbf{v} : \mathbf{v} \text{ satisfy constraints (5) – (12)}\}, \\ \Omega_{\mathbf{v}}^{\text{UE}} &= \{\mathbf{v} : \mathbf{v} \text{ satisfy constraints (13) – (22)}\}.\end{aligned}\quad (23)$$

We can formulate the relaxed problem (RP) for the network design with congestion pricing when the traffic assignment follows the system optimal (SO) principle as follows. This enables us to obtain the optimal solution \mathbf{x}^* with SO traffic assignment.

[RP]

$$\begin{aligned}\min_{(\mathbf{x},\mathbf{0}) \in \Omega_{\mathbf{x},\boldsymbol{\tau}}, \mathbf{v} \in \Omega_{\mathbf{v}}^{\text{RP}}} & \sum_{a \in A \setminus \bar{A}} t_a(v_a, c_a) v_a + \sum_{a \in \bar{A}} x_a t_a(v_a, c_a^1) v_a \\ & + \sum_{a \in \bar{A}} (1 - x_a) t_a(v_a, c_a^2) v_a\end{aligned}\quad (24)$$

In the relaxation-based method, we solve [RP] by iteratively excluding the solution \mathbf{x}^* . We do so by adding the following constraint to [RP] iteratively:

$$\begin{aligned}\sum_{\text{all } a \in \bar{A} \text{ that satisfying } x_a^* = 0} x_a \\ + \sum_{\text{all } a \in \bar{A} \text{ that satisfying } x_a^* = 1} (1 - x_a) \geq 1\end{aligned}\quad (25)$$

Here, we define the pricing problem (PP) when the network design variable \mathbf{x} is given as \mathbf{x}^* .

[PP]

$$\begin{aligned}\min_{(\mathbf{x}^*, \boldsymbol{\tau}) \in \Omega_{\mathbf{x},\boldsymbol{\tau}}, \mathbf{v} \in \Omega_{\mathbf{v}}^{\text{UE}}} & \sum_{a \in A \setminus \bar{A}} t_a(v_a, c_a) v_a + \sum_{a \in \bar{A}} x_a t_a(v_a, c_a^1) v_a \\ & + \sum_{a \in \bar{A}} (1 - x_a) t_a(v_a, c_a^2) v_a\end{aligned}\quad (26)$$

The relaxation-based method is as follows:

Step 0. Define a set $\bar{\Omega}_{\mathbf{x}} = \emptyset$ that contains all generated solutions for generating the cuts given by constraint (25). Define the upper bound $UB = +\infty$, incumbent optimal solution $(\mathbf{x}^{\text{opt}}, \boldsymbol{\tau}^{\text{opt}})$.

Step 1. Reformulate [RP] as [RP-1], as follows:

[RP-1]

$$\begin{aligned}\min_{(\mathbf{x},\mathbf{0}) \in \Omega_{\mathbf{x},\boldsymbol{\tau}}, \mathbf{v} \in \Omega_{\mathbf{v}}^{\text{RP}}} & \sum_{a \in A \setminus \bar{A}} t_a(v_a, c_a) v_a + \sum_{a \in \bar{A}} t_a(v_a^1, c_a^1) v_a^1 \\ & + \sum_{a \in \bar{A}} t_a(v_a^2, c_a^2) v_a^2\end{aligned}\quad (28)$$

s.t.

$$\begin{aligned}v_a &= v_a^1 + v_a^2, \quad \forall a \in \bar{A}, \\ 0 &\leq v_a^1 \leq Lx_a, \quad \forall a \in \bar{A}, \\ 0 &\leq v_a^2 \leq L(1 - x_a), \quad \forall a \in \bar{A}, \\ 0 &\leq \tau_a \leq Lx_a, \quad \forall a \in \bar{A}, \\ x_a &\in \{0, 1\}, \quad \forall a \in \bar{A},\end{aligned}\quad (28)$$

where v_a^1 and v_a^2 are auxiliary variables. Solve [RP-1] with the following constraints using the outer-approximation algorithm presented in Wang et al. [33]:

$$\sum_{a \in \bar{A}} (1 - \bar{x}_a) x_a + \sum_{a \in \bar{A}} \bar{x}_a (1 - x_a) \geq 1, \quad \forall \bar{\mathbf{x}} \in \bar{\Omega}_{\mathbf{x}}. \quad (29)$$

If the problem is infeasible, we have enumerated all feasible solutions and hence $(\mathbf{x}^{\text{opt}}, \boldsymbol{\tau}^{\text{opt}})$ is the optimal solution and we stop. Otherwise, obtain the provisional optimal design denoted by \mathbf{x}^* and the provisional optimal value denoted by Obj_{RP} of [RP] under constraint (29). If $UB \leq Obj_{\text{RP}}$, $(\mathbf{x}^{\text{opt}}, \boldsymbol{\tau}^{\text{opt}})$ is the optimal solution, stop. Otherwise, go to step 2.

Step 2. Solve [PP] by fixing \mathbf{x} at \mathbf{x}^* and apply the relaxation algorithm presented in Wang et al. [27]. Then, we can obtain the optimal value Obj_{PP} , optimal pricing $\boldsymbol{\tau}^*$, and vehicle flow \mathbf{v}^* . If $UB \leq Obj_{\text{PP}}$, then $(\mathbf{x}^*, \boldsymbol{\tau}^*)$ is not better than $(\mathbf{x}^{\text{opt}}, \boldsymbol{\tau}^{\text{opt}})$ and, hence, we set $\bar{\Omega}_{\mathbf{x}} := \bar{\Omega}_{\mathbf{x}} \cup \{\mathbf{x}^*\}$ and go to step 1. Else, set $UB = Obj_{\text{PP}}$ and $(\mathbf{x}^{\text{opt}}, \boldsymbol{\tau}^{\text{opt}}) = (\mathbf{x}^*, \boldsymbol{\tau}^*)$. If $UB \leq Obj_{\text{RP}}$ and $(\mathbf{x}^{\text{opt}}, \boldsymbol{\tau}^{\text{opt}})$ is the optimal solution, stop. Otherwise, set $\bar{\Omega}_{\mathbf{x}} := \bar{\Omega}_{\mathbf{x}} \cup \{\mathbf{x}^*\}$ and go to step 1.

4. Numerical Tests

4.1. Network Description. In the following sections, to evaluate the effectiveness of the proposed model and solution method, we present some numerical examples for a nine-node network with 18 links [37]. Suppose that the travel-time cost functions of all links are defined as having the following BPR form:

$$t_a(v_a, C_a) = t_a^0 \left(1 + 0.15 \left(\frac{v_a}{C_a} \right)^4 \right), \quad (30)$$

where t_a^0 is the free-flow travel-time cost on link a and C_a is the capacity of link a . Figure 2 shows the initial network without converting regular links to AV links. The tuple above

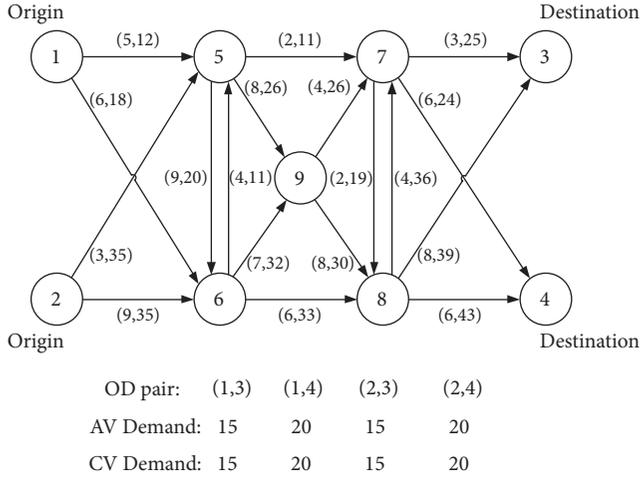


FIGURE 2: Nine-node network.

each link a denotes the free-flow travel-time cost and initial capacity when all links are regular. There are four OD pairs in the network, and the OD demands of all the travel modes are shown at the bottom of Figure 2.

In the initial network, all links are regular links and all vehicles are allowed to operate on these links without congestion pricing. We can obtain vehicle flows in the initial network by solving UE problems (4)-(12); when setting candidate AV link set $\tilde{A} = \emptyset$ without congestion pricing. After solving this UE problem for vehicle flows, we find that no AVs run on links (5,6), (6,5), (6,9), (7,8), (8,7), and (8,3) in the initial network. So we exclude these links from the candidate link set and form the following candidate link set:

$$\tilde{A} = \{(1,5), (1,6), (2,5), (2,6), (5,7), (5,9), (6,8), (9,7), (9,8), (7,3), (7,4), (8,4)\}. \quad (31)$$

If any link $a \in \tilde{A}$ is converted to an AV link, then the capacity of link a is tripled and only AVs are allowed to run on it. For any link $a \in \tilde{A}$ that is not converted to an AV link, the road planner can implement congestion pricing for CVs.

In the following tests, we used a personal computer with an Intel Core (TM) i7 4700MQ CPU, 16GB RAM, and Windows 7 Professional operating system. We coded the proposed solution method using Python and called Gurobi 7.5.2 and Ipopt 3.12.7 to solve the NDP with congestion pricing. We ran the codes for all cases in the following tests and the longest computation time was less than five minutes.

4.2. Convergence of Relaxation-Based Method. We tested the convergence of the relaxation-based method, and Figure 3 shows the lower and upper bounds at each iteration of this method. We can see that the gap between lower and upper bounds decreases sharply and finally converges to zero. Table 1 shows the optimal network design with congestion pricing for this case. Table 2 shows the link flows in the network when the optimal network design with congestion pricing is implemented.

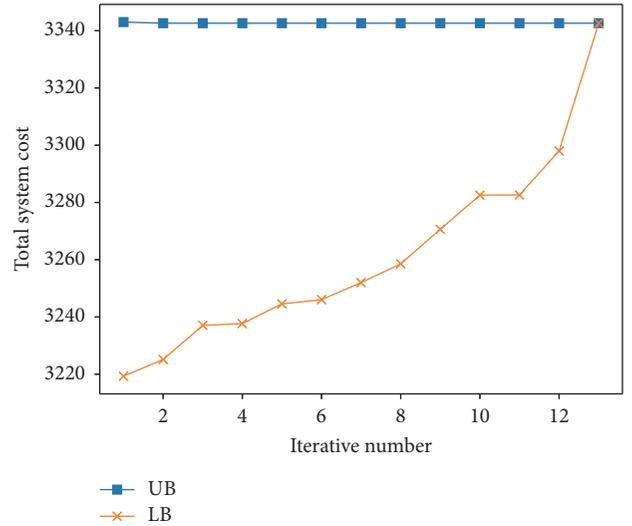


FIGURE 3: Lower and upper bounds at each iteration of the relaxation-based method.

TABLE 1: Optimal network design with congestion pricing.

Candidate link	x_a	τ_a
(1,5)	1	0.000
(1,6)	0	4.756
(2,5)	0	3.272
(2,6)	0	5.821
(5,7)	1	0.000
(5,9)	0	3.350
(6,8)	0	8.732
(9,7)	0	3.132
(9,8)	0	1.532
(7,3)	1	0.000
(7,4)	1	0.000
(8,4)	0	4.593

4.3. Model Comparison. In this section, we use four models to demonstrate the efficiency of the NDP with congestion pricing. The first model is to do nothing, in that we set no links for AVs and no tolls for CVs. The second model is a pricing scheme for CVs without any network design. The third model is an NDP without congestion pricing. The fourth model is the NDP with congestion pricing, as proposed. The total travel-time costs associated with these four models are 5886.276, 5440.112, 3507.934, and 3342.627, respectively. We can clearly see that using both network design and congestion pricing can improve the traffic conditions with heterogeneous AV and CV travel modes. In addition, the road planner can implement these methods simultaneously, which realizes better results than implementing the network design or congestion pricing separately.

4.4. Sensitivity Analyses. Since many parameters can impact the total travel-time cost in a network, here, we perform sensitivity analyses. All tests in this section are based on the network proposed in Section 4.1.

TABLE 2: Link flows in the network upon implementation of the optimal network design with congestion pricing.

Link	AV link flow	CV link flow
(1,5)	35.000	0.000
(1,6)	0.000	35.000
(2,5)	35.000	0.000
(2,6)	0.000	35.000
(5,6)	0.000	0.000
(6,5)	0.000	0.000
(5,7)	55.774	0.000
(5,9)	14.226	0.000
(6,8)	0.000	42.768
(6,9)	0.000	27.232
(9,7)	4.450	18.119
(9,8)	9.775	9.113
(7,8)	0.000	18.119
(8,7)	0.000	0.000
(7,3)	30.000	0.000
(7,4)	30.225	0.000
(8,3)	0.000	30.000
(8,4)	9.775	40.000

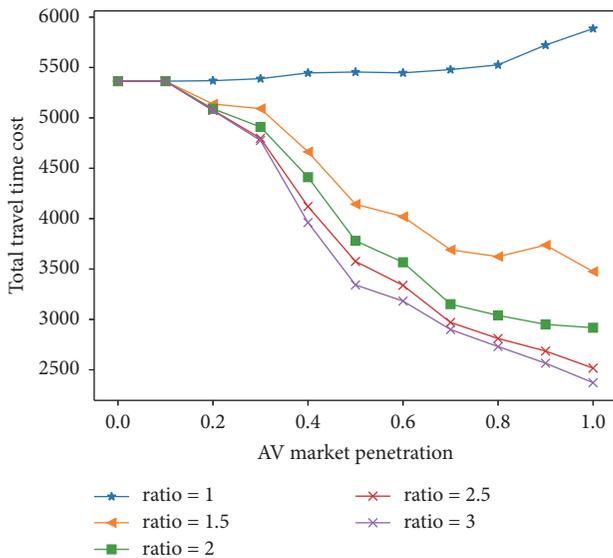


FIGURE 4: Variances of total travel-time cost with AV market penetrations.

Figure 4 shows variances in the total travel-time cost against AV market penetrations. “ratio = 3” indicates that the AV link capacity is tripled when it is converted from a regular to an AV link. We can see that a greater adoption of AVs always leads to less total travel-time cost for ratio = 2, 2.5, 3. However, this is not true for ratio = 1, 1.5. This means that AVs should be promoted only when the AV technique is mature and AV link capacity can be increased significantly; otherwise, AV promotion may cause the opposite effect.

When the AV technique matures, more people will choose the AV mode to benefit from its higher capacity and

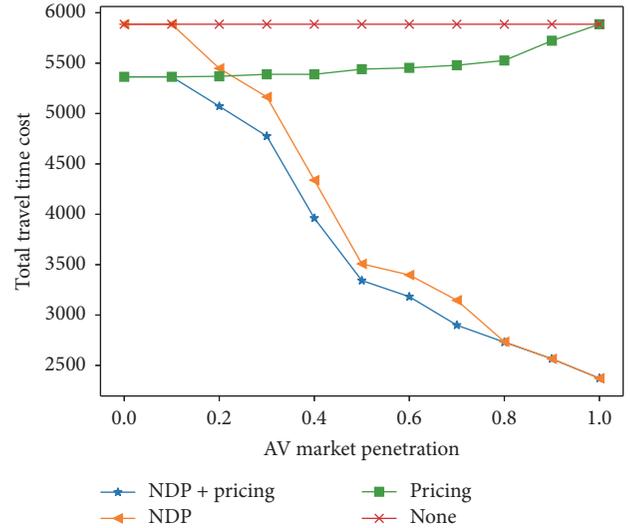
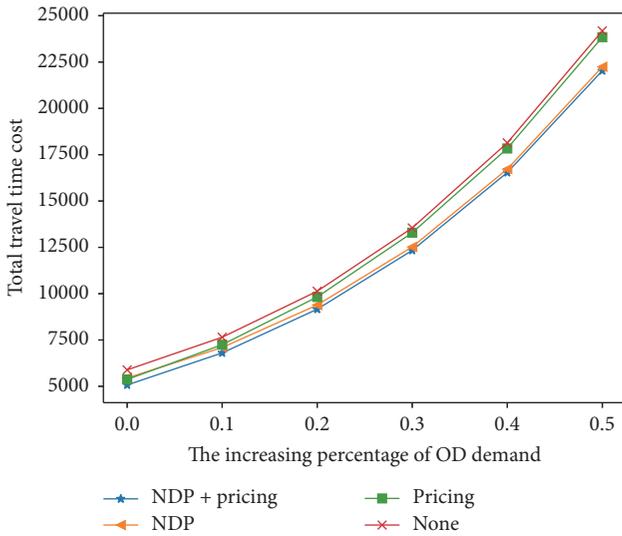


FIGURE 5: Four models at different AV market penetrations.

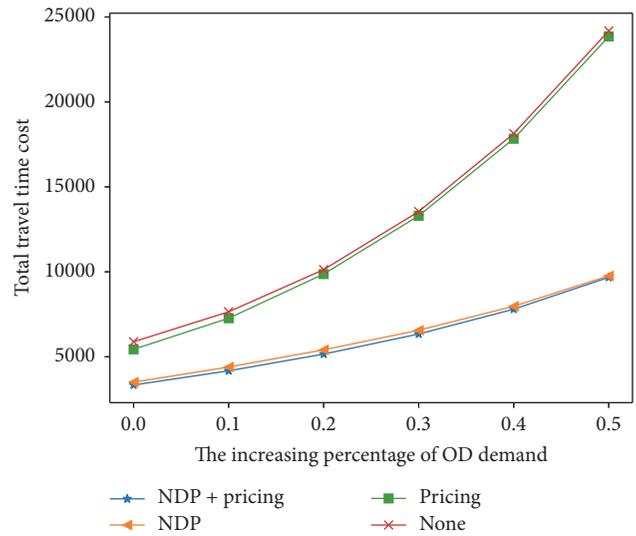
lower travel time. By fixing the total demand of the network, we tested the models described in Section 4.3 at different AV market penetrations. Figure 5 shows a comparison of the four models, in which we can see that the NDP with congestion pricing is always the most effective. Also, we can see that congestion pricing is effective when the AV market penetration is low and network design is effective when AV penetration is high. The reason for this is clear: when AV market penetration is low, congestion pricing to control CVs is effective. When AV penetration is high, a network design for AVs is effective. This means that road planners should concentrate on different strategies for different AV market penetrations. In addition, we find that the NDP with congestion pricing strongly outperforms either NDP or pricing when both AVs and CVs have large market penetrations. This shows that the proposed NDP with congestion pricing can be an effective method for alleviating traffic congestion in the transitional period characterized by heterogeneous AV and CV travel modes.

Assuming that the initial demand is that shown in Figure 2, Figures 6(a)–6(c) show variances in the total travel-time costs for the four models with increasing percentages of OD demand and AV market penetration. Given a fixed AV market penetration, we can see that the effect of pure congestion pricing becomes worse with increasing demand. However, the effect of a pure network design improves with increasing demand. The gap between the total travel-time cost by the NDP with congestion pricing and that by an NDP alone is almost the same even if the OD demand increases. Also, we can see that the NDP becomes more effective as OD demand increases, but congestion pricing does not. As such, we can infer that AV promotion and AV link design may be the best solution for alleviating traffic congestion as traffic demand continues to grow.

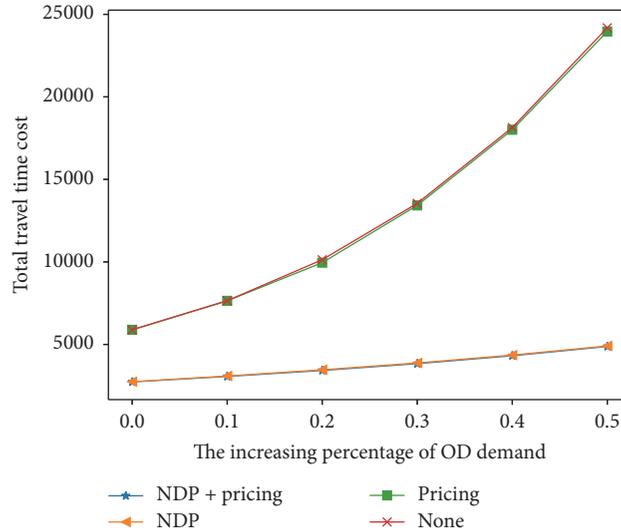
Since high congestion pricing for CVs would generate public resentment, we set an upper bound for pricing τ_a . Figure 7 shows the variance of total travel-time cost with



(a) Variances of total travel-time cost with increasing percentage of OD demand when AV market penetration = 20%



(b) Variances of total travel-time cost with increasing percentage of OD demand when AV market penetration = 50%



(c) Variances of total travel-time cost with increasing percentage of OD demand when AV market penetration = 80%

FIGURE 6

respect to the upper bound for congestion pricing. Clearly, setting a reasonable upper bound for congestion pricing cannot only reduce public discontent, but also can effectively reduce total travel-time cost.

5. Conclusions and Future Research

In this paper, we proposed a bi-level network design model comprising AV links and congestion pricing to alleviate traffic congestion when AVs and CVs are both operating in a transportation network. In this model, the upper-level road planner chooses the optimal network design for AV links and congestion pricing while lower-level travelers choose their routes based on their individual travel costs and thereby

achieve user equilibrium. We conducted numerical tests on a nine-node network and the results reveal the feasibility of the proposed model and solution method. Our main findings are shown as follows:

- (1) Both network design and congestion pricing can alleviate traffic congestion. The integrated optimization of a network design and congestion pricing can achieve better traffic conditions than either a single network design or a congestion pricing.
- (2) If we compare the performance outcome of pure congestion pricing with that of a network design, we can see that congestion pricing is effective when AV market penetration is low and a network design is effective when AV penetration is high. As such, the

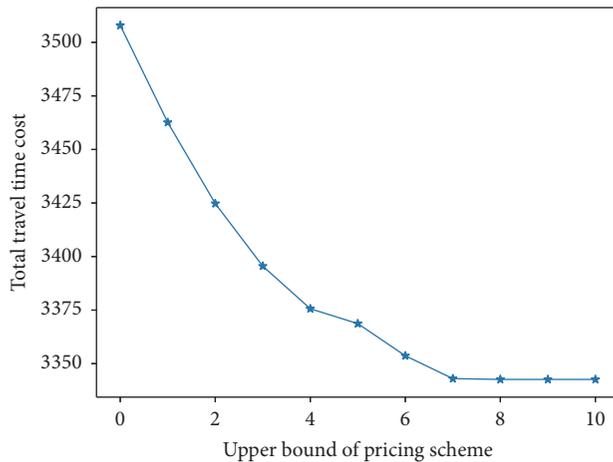


FIGURE 7: Variance of total travel-time cost with respect to the upper bound of congestion pricing.

road planner should employ different strategies for different AV market penetrations.

- (3) Since network design becomes more effective as OD demand increases, it follows that AV promotion and AV link design may be the best solution for alleviating traffic congestion when traffic demand becomes increasingly large.
- (4) In the transitional period characterized by heterogeneous AV and CV travel modes, the NDP with congestion pricing strongly outperforms either network design or congestion pricing alone when both AVs and CVs have fairly large market penetrations. In addition, if we add an appropriate upper bound for congestion pricing, we can reduce public discontent without losing the effectiveness of the NDP with congestion pricing.

Based on our proposed model and results, we suggest the consideration of future studies as follows:

- (1) We suggest NDP with a tradable credit scheme to alleviate traffic congestion. Although the proposed NDP with congestion pricing can alleviate traffic congestion and an appropriate upper bound for congestion pricing can reduce public discontent without losing its effectiveness, congestion pricing can be viewed as unfair or as a flat tax [38]. If we consider the NDP with a tradable credit scheme, then the problem of unfairness is fully resolved.
- (2) Multiclasss of travelers can be considered. We can infer the relationship between the classes of travelers and the perceived value of their travel time, based on the population income distribution. This could be helpful in tackling real-world situations and yield interesting findings.
- (3) We suggest optimal design with congestion pricing considering variational AV market penetration. In this paper, we considered AV market penetration

to be exogenous. However, AV market penetration can be viewed as an endogenous variable and the evolution of this penetration can be forecast using prediction models, e.g., the diffusion model [13, 39]. Based on this assumption, a time-dependent NDP with congestion pricing could be formulated to alleviate traffic congestion over a long period.

- (4) The optimal structure of intermodal transportation network is very important for both passenger traffic and logistics [40]; network design of intermodal transportation network with AVs might be an important research topic in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank Professor Xiaoning Zhang of Tongji University for his time in proofreading the paper. This study has been substantially supported by the research grants from the National Natural Science Foundation Council of China (nos. 71531011, 71601142, and 71471134), Shanghai Shuguang Program (13SG23), and Shanghai Pujiang Program (16PJC090).

References

- [1] H. Wang, W. H. K. Lam, X. Zhang, and H. Shao, "Sustainable transportation network design with stochastic demands and chance constraints," *International Journal of Sustainable Transportation*, vol. 9, no. 2, pp. 126–144, 2015.
- [2] H. Wang and X. Zhang, "Game theoretical transportation network design among multiple regions," *Annals of Operations Research*, vol. 249, no. 1-2, pp. 97–117, 2017.
- [3] R. Z. Farahani, E. Miandoabchi, W. Y. Szeto, and H. Rashidi, "A review of urban transportation network design problems," *European Journal of Operational Research*, vol. 229, no. 2, pp. 281–302, 2013.
- [4] H. Yang and M. G. H. Bell, "Models and algorithms for road network design: a review and some new developments," *Transport Reviews*, vol. 18, no. 3, pp. 257–278, 1998.
- [5] Q. Meng, Z. Liu, and S. Wang, "Optimal distance tolls under congestion pricing and continuously distributed value of time," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 5, pp. 937–957, 2012.
- [6] Z. Liu, Y. Zhang, S. Wang, and Z. Li, "A trial-and-error method with autonomous vehicle-to-infrastructure traffic counts for cordon-based congestion pricing," *Journal of Advanced Transportation*, vol. 2017, Article ID 9243039, pp. 1–8, 2017.

- [7] H. Wang and X. Zhang, "Joint implementation of tradable credit and road pricing in public-private partnership networks considering mixed equilibrium behaviors," *Transportation Research Part E: Logistics and Transportation Review*, vol. 94, pp. 158–170, 2016.
- [8] S. Le Vine, A. Zolfaghari, and J. Polak, "Autonomous cars: The tension between occupant experience and intersection capacity," *Transportation Research Part C: Emerging Technologies*, vol. 52, pp. 1–14, 2015.
- [9] N. Jiang, "Optimal signal design for mixed equilibrium networks with autonomous and regular vehicles," *Journal of Advanced Transportation*, vol. 2017, pp. 1–13, 2017.
- [10] J. Wang, X. Zhang, and H. M. Zhang, "Parking permits management and optimal parking supply considering traffic emission cost," *Transportation Research Part D: Transport and Environment*, 2016.
- [11] F. Zhu and S. V. Ukkusuri, "A linear programming formulation for autonomous intersection control within a dynamic traffic assignment and connected vehicle environment," *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 363–378, 2015.
- [12] P. Tientrakool, Y.-C. Ho, and N. F. Maxemchuk, "Highway capacity benefits from using vehicle-to-vehicle communication and sensors for collision avoidance," in *Proceedings of the IEEE 74th Vehicular Technology Conference, VTC Fall*, The Hilton San Francisco Union Square, San Francisco, USA, September 2011.
- [13] Z. Chen, F. He, L. Zhang, and Y. Yin, "Optimal deployment of autonomous vehicle lanes with endogenous market penetration," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 143–156, 2016.
- [14] X. Zhang and H. Yang, "The optimal cordon-based network congestion pricing problem," *Transportation Research Part B: Methodological*, vol. 38, no. 6, pp. 517–537, 2004.
- [15] X. Zhang, H. M. Zhang, H.-J. Huang, L. J. Sun, and T.-Q. Tang, "Competitive, cooperative and stackelberg congestion pricing for multiple regions in transportation networks," *Transportmetrica*, vol. 7, no. 4, pp. 297–320, 2011.
- [16] H. Yang and H. J. Huang, *Mathematical and Economic Theory of Road Pricing*, Elsevier, Oxford, UK, 2005.
- [17] Y. Ji, M. Xu, H. Wang, and C. Tan, "Commute equilibrium for mixed networks with autonomous vehicles and traditional vehicles," *Journal of Advanced Transportation*, vol. 2017, Article ID 6218363, pp. 1–10, 2017.
- [18] S. Deb, L. Strawderman, D. W. Carruth, J. DuBien, B. Smith, and T. M. Garrison, "Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 178–195, 2017.
- [19] C. Häne, T. Sattler, and M. Pollefeys, "Obstacle detection for self-driving cars using only monocular cameras and wheel odometry," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2015*, pp. 5101–5108, Germany, October 2015.
- [20] A. Aparicio, L. Sanz, G. Burnett et al., *Advancing Active Safety Towards the Protection of Vulnerable Road Users: The Prospect Project*, 2017.
- [21] M. W. Levin and S. D. Boyles, "A multiclass cell transmission model for shared human and autonomous vehicle roads," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 103–116, 2016.
- [22] P. Godsmark and G. Kakkar, "Why automated vehicle zones make sense," *The Canadian Automated Vehicles Centre of Excellence*, 2014.
- [23] Z. Chen, F. He, Y. Yin, and Y. Du, "Optimal design of autonomous vehicle zones in transportation networks," *Transportation Research Part B: Methodological*, vol. 99, pp. 44–61, 2017.
- [24] N. Jiang, X. Zhang, and H. Wang, "Simultaneous optimization of road tolls and tradable credits in public-private mixed networks," *PROMET - Traffic and Transportation*, vol. 29, no. 6, p. 603, 2017.
- [25] A. Koh, S. Shepherd, and A. Sumalee, "Second best toll and capacity optimisation in networks: Solution algorithm and policy implications," *Transportation*, vol. 36, no. 2, pp. 147–165, 2009.
- [26] X. Zhang and B. van Wee, "Enhancing transportation network capacity by congestion pricing with simultaneous toll location and toll level optimization," *Engineering Optimization*, vol. 44, no. 4, pp. 477–488, 2012.
- [27] G. Wang, Z. Gao, M. Xu, and H. Sun, "Joint link-based credit charging and road capacity improvement in continuous network design problem," *Transportation Research Part A: Policy and Practice*, vol. 67, pp. 1–14, 2014.
- [28] M. Xu, G. Wang, S. Grant-Muller, and Z. Gao, "Joint road toll pricing and capacity development in discrete transport network design problem," *Transportation*, vol. 44, no. 4, pp. 731–752, 2017.
- [29] Y. Lim, B. G. Heydecker, and S. Lee, "A continuous network design model in stochastic user equilibrium based on sensitivity analysis," *Journal of Advanced Transportation*, vol. 39, no. 1, pp. 63–79, 2005.
- [30] G. M. Wang, Z. Y. Gao, and M. Xu, "An MPEC formulation and its cutting constraint algorithm for continuous network design problem with multi-user classes," *Applied Mathematical Modelling*, vol. 38, no. 5–6, pp. 1846–1858, 2014.
- [31] D. Z. W. Wang and H. K. Lo, "Global optimum of the linearized network design problem with equilibrium flows," *Transportation Research Part B: Methodological*, vol. 44, no. 4, pp. 482–492, 2010.
- [32] H. Liu and D. Z. W. Wang, "Global optimization method for network design problem with stochastic user equilibrium," *Transportation Research Part B: Methodological*, vol. 72, pp. 20–39, 2015.
- [33] S. Wang, Q. Meng, and H. Yang, "Global optimization methods for the discrete network design problem," *Transportation Research Part B: Methodological*, vol. 50, pp. 42–60, 2013.
- [34] H. Sun, J. Wu, W. Wang, and Z. Gao, "Reliability-based traffic network design with advanced traveler information systems," *Information Sciences*, vol. 287, pp. 121–130, 2014.
- [35] Y. Kong, S. Le Vine, and X. Liu, "Capacity impacts and optimal geometry of automated cars' surface parking facilities," *Journal of Advanced Transportation*, vol. 2018, pp. 1–13, 2018.
- [36] J. F. Bard, "Some properties of the bilevel linear programming," *Journal of Optimization Theory and Applications*, vol. 2, pp. 146–164, 1991.
- [37] D. W. Hearn and M. V. Ramana, "Solving congestion toll pricing models," in *Equilibrium and Advanced Transportation Modelling*, pp. 109–124, Springer, 1998.
- [38] H. Yang and X. Wang, "Managing network mobility with tradable credits," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 580–594, 2011.

- [39] H. Yang and Q. Meng, "Modeling user adoption of advanced traveler information systems: Dynamic evolution and stationary equilibrium," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 10, pp. 895–912, 2001.
- [40] H. Wang, X. Wang, and X. Zhang, "Dynamic resource allocation for intermodal freight transportation with network effects: Approximations and algorithms," *Transportation Research Part B: Methodological*, vol. 99, pp. 83–112, 2017.

Research Article

Minimizing the Average Delay at Intersections via Presignals and Speed Control

Mina Ghanbarikarekani ¹, Xiaobo Qu ², Michelle Zeibots,³ and Weiwei Qi ⁴

¹*School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia*

²*Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden*

³*Institute for Sustainable Futures, University of Technology Sydney, Sydney, NSW 2007, Australia*

⁴*School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China*

Correspondence should be addressed to Weiwei Qi; qwwhit@163.com

Received 27 January 2018; Accepted 8 April 2018; Published 29 May 2018

Academic Editor: David Z. W. Wang

Copyright © 2018 Mina Ghanbarikarekani 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.

Urban intersections have been well recognized as bottlenecks of urban transport systems. It is thus important to propose and implement strategies for increasing the efficiency of public and private transportation systems as a whole. In order to achieve this goal, an additional signal could be set up near the intersection to give priority to buses through stopping vehicles in advance of the main intersection as a presignal. It has been increasingly popular in urban cities. While presignals indeed reduce the average delay per traveler, they cause extra stops of private vehicles, which might compromise the overall efficiency, safety, and sustainability. This paper aims to propose a model to improve presignals by reducing the vehicles' number of stops behind the presignals. By applying the method, vehicles would be able to adjust their speed based on traffic conditions as well as buses' speed and approach. Numerical analyses have been conducted to determine the conditions required for implementing this method.

1. Introduction

Traffic congestion is one of the most significant issues in big cities that dissatisfy inhabitants. This matter has remarkably inevitable impacts on people's daily life in terms of travel time and health that resulted from increasing delay, traffic congestion, and air pollution [1–4]. Therefore, promising methodologies, which mitigate traffic congestion, play an important role in improving the life situation and transportation systems in main cities, specifically in their central business district [5]. Public transportation plays a vital role in decreasing traffic congestion as it has higher passenger factor than private vehicles, while it occupies less road space [6]. Hence, one of the most efficient ways that results in calming traffic congestion is developing public transit and encouraging people to use it rather than their private vehicles [7]. To serve this purpose, it is necessary to enhance the performance of public transportation systems in terms of delay and travel time.

Traffic control strategies often prioritize public transportation to motivate people not to use their private vehicles

[8–10]. Presignaling for buses has been proven one of the effective strategies to reduce the average delay per passenger [11]. Presignal is a kind of traffic signal that is implemented in advance of main signalized intersections and provides priority for buses in choosing their approach behind the main intersections by giving red signal to private vehicles. Although this method has remarkably efficient effects on reducing buses' delay and travel time and increasing their speed, it has several disadvantages in traffic flow. More specifically, presignals impose additional stops to private cars that lead to the increase of their delay and travel time. In other words, using presignals enhances the functionality of buses through prioritizing them at intersections, while it does not consider private vehicles.

This study aims to give priority to buses as well as improve the performance of the private vehicles through minimizing their number of stops while reaching presignals. For the sake of alleviating the cars' number of stops, it is suggested to amend their speed based on traffic signal timing, buses approach, and traffic condition ahead. Consequently,

the traffic flow of the private cars will be continuous and balanced rather than being discrete.

The rest of the paper is structured as follows. Section 2 explains background studies carried out related to the topic. Section 3 describes the methodology, the essential assumptions, and its procedure. Numerical analysis of the model, the results, and discussion are presented in Section 4. Section 5 shows the conclusion.

2. Literature Review

Presignalizing is one of the most recent innovations that have been paid attention over the last two decades. The bus presignal was defined by Peake [12] as the traffic signals implemented behind intersections for managing traffic flow and giving buses priority. The first time using presignals for prioritizing buses was proposed by Wu and Hounsell [13]. In their research, three different categories were presented including (1) implementing presignals with buses uncontrolled, (2) presignals with controlling both buses and cars, and finally (3) giving the red signal to private vehicles during arriving buses to presignal and then giving the red signal to the bus lane. It needs to be mentioned that, according to studies conducted by Wu and Hounsell [13], Kumara and Hounsell [14], He et al. [15], Xuan [16], and Xuan et al. [17], using and implementing presignals upstream of the urban intersections lead to the decrease of the discharge rate of the intersections and the waste of the intersection's green time. In other words, fewer vehicles can be discharged from presignals than the capacity and green time of the main intersection.

Kumara and Hounsell [14] proposed queue relocation to avoid wasting the main intersections' green time. Two priority methods in presignalized intersections, queue relocation and bus priority, were suggested in their research. Queue relocation keeps private vehicles at presignal stop line due to saving the green time of the main intersection. Moreover, bus priority is supplied by detectors embedded in vehicles. They indicated that presignals play a remarkable role in prioritizing buses in oversaturated intersections throughout queue relocation and bus priority.

Another solution offered to solve the problem of intersections' discharge rate is using mid-block presignals posed by Xuan et al. [17]. They demonstrated that this solution causes traffic flow to be stored efficiently between presignals and main intersections. Moreover, for increasing the intersections' discharge rate, an adaptive algorithm was proposed by He et al. [15]. This algorithm can control presignals regarding real-time demand for private and public transportation. It was achieved by accurate statistics and real-time detection. They indicated that presignals with adaptive control algorithm lead to the stimulation of the buses' use and reduction of person delay more than continuous and interrupted bus lane strategies, respectively. Increasing the intersections flow capacity is one of the significant parameters that affects the intersections efficiency fundamentally. Another considerable study carried out for increasing the intersections' capacity is Xuan's research [16]. Xuan has proposed a method in order to raise the signalized intersections flow capacity using tandem

design. In this study, left-turning and through-moving vehicles were sorted throughout a mid-block presignal. Xuan conducted studies on the requirements of the blocks' length for reaching the optimal capacity and the effects of the tandem design on minimizing these requirements. It was shown that this tandem design had increased the flow capacity of cars and buses in intersections equipped with presignals.

In addition to the low discharge rate of the intersections, Wu and Hounsell [13] have explained another issue of presignals in their research and that is an extension of queue length to the upstream intersection. They have suggested vital assumptions for estimating traffic signal timing to avoid these mentioned issues.

Kejun [18] has conducted a study about prioritizing buses at a single intersection using presignal and passive priority through introducing Bus Advance Area between stop lines of the presignalized and main intersections. In addition, Kejun investigated the efficiency of presignal by simulating the studied intersection in VISSIM; consequently, it was demonstrated that although presignals increase the efficiency of the buses, they have a negative impact on the private vehicles' performance because of their additional stop behind presignals. For eliminating this impact at intersections equipped with presignal, He et al. [19] have suggested a control algorithm with an online performance. This algorithm has been modelled in microsimulation software, VISSIM. In their research, they implemented the algorithm with the bus lane, mixed lane, and presignal strategies. In the end, by comparing these methods, it was demonstrated that using presignals with the proposed algorithm not only prioritizes buses but also maintains proficiency of the private vehicles. In addition to He et al.'s studies, Guler and Menendez have played a significant role in improving the functionality of private vehicles as well as bus priority. Guler and Menendez [20] have estimated delays of cars and buses in presignalized intersections analytically using queuing theory. In their research, they have computed traveler delays in implemented presignals and allocated a lane to buses. By comparing the commuters' delay, it was concluded that presignal systems minimize the delay more than dedicated bus lanes. Consequently, implementing presignals has fundamental effects on buses due to their high capacities. In addition, Guler and Menendez [21] have presented a practical instruction about using presignals upstream of the intersections. The influences of implementing presignal on intersections comparing to other bus priority strategies pave the way to determine the conditions of applying presignals in arterials. It should be mentioned that they have proposed the presignals instruction to improve transit services and private transportation systems simultaneously. Furthermore, Guler et al. [22] have suggested an innovative strategy to prioritize buses as well as improving the functionality of cars. This strategy provides dynamic timing, for presignals, which can be activated or deactivated depending on the traffic situation. In their research, it has been considered that implementing presignals leads to increasing cars' delay in undersaturated intersections and reducing their discharging rate in oversaturated intersections. However, using the single-lane strategy can mitigate the mentioned undesirability.

3. Methodology

In this section, we first present some essential assumptions for the model. We further describe the model and its procedures in Section 3.2. In addition, the model's objective function based on the initial queue delay and the required parameters for the model are elaborated.

3.1. The Model Assumptions. In this study, there are some assumptions notified for proposing and presenting a model:

- (i) Without loss of generality, we assume there are two lanes.
- (ii) The intersection is controlled by a fixed-time traffic signal and equipped with presignal.
- (iii) It is assumed that the studied intersection is isolated; that is, the performance of the adjacent intersections has no influence on it.
- (iv) A Variable Message Sign (VMS) is installed for alarming the private cars.
- (v) The Automatic Vehicle Location system (AVL) is utilized in each bus for the sake of determining the buses' location, their approach, and selected lane.
- (vi) Detectors are needed throughout the area behind the stop line where buses can change their lane. These detectors also declare the existence of other cars and the bus's selected lane.
- (vii) In the point VMS is installed, another detector is needed to count the private cars that occupied the distance between the VMS and the area behind the stop line.
- (viii) The movement of the buses and cars is assumed static, and their speed value is constant, which is based on road rules.

3.2. The Model Procedure. The study aims to minimize the number of stops for optimizing the performance of the buses and private vehicles in urban arterials with signalized intersections equipped with presignals. Therefore, it is essential that

- (i) buses be located in the first line of the queue behind the stop line; presignals are the appropriate methods for meeting this requirement by prioritizing buses behind intersections;
- (ii) cars be at a moderate speed, without any stop or with the minimum stops behind presignals. To this end, it is required to minimize the initial queue behind presignals. More specifically, it would be necessary to minimize the initial queue delay [23] at presignals through the objective function proposed in (1).

Objective function:

$$F(x) = \min [d_3] \quad (1)$$

subject to

$$d_3 = \frac{1800 \times Q_b \times (1 + u) \times t}{CT} \quad (2)$$

If $v_{\text{car,suggested}} \geq 5 \text{ km/h}$
then $Q_b = 0$,

where

T is the duration of analysis period (h),

Q_b is the initial queue at the start of period T (veh),

C is the adjusted lane group capacity (veh/h),

t is the duration of unmet demand in T (h), ($t = 0$ if $Q_b = 0$, otherwise $t = \min[T, Q_b/C[1 - \min(1, V/C)]]$ if $Q_b \neq 0$),

u is the delay parameter ($u = 0$ if $t < T$, otherwise $u = 1 - CT/Q_b[1 - \min(1, V/C)]$ if $t \geq T$),

$v_{\text{car,suggested}}$ is the suggested speed of the car in arterials using the model (km/h).

It is required to identify some parameters in order to estimate the suggested speed of cars ($v_{\text{car,suggested}}$). These parameters would be gained through the following three steps.

Step 1 (estimate the optimal distance for implementing presignals). At this distance, buses could select their desirable lane at the main intersection. Additionally, the AVL applied in buses declare their real-time position to private vehicles in order to reduce their speeds due to prioritizing the buses and their selected lane (shown in Figure 1 as d_{bus}).

There are several factors that need to be considered for the sake of estimating this distance. These factors are based on buses movements for changing their lane and reaching the stop line of the main intersection. In this paper, the distance of implementing presignals (d_{bus}) is obtained by modifying the distance proposed by Guler and Menendez [21]. They assumed that bus stopped at the presignal so its initial speed to move towards the main intersection was zero. However, in this research, we suppose that buses do not stop at presignals and they reach the main intersection with an initial speed. Moreover, it is assumed that buses change their lane and arrive at the main intersection when traffic signal starts at red. Therefore, d_{bus} can be calculated by the following equation:

$$d_{\text{bus}} = \left[(c - r_{ms})_1 \times \left(\frac{C}{k_{\text{jam}}} - v_1 \right) + (c - r_{ms})_2 \right] \times \left(0 - \frac{C}{k_{\text{jam}}} \right) \times 1000, \quad (3)$$

where

d_{bus} is the distance between the presignal and the main intersection (m),

c is the common cycle length of the traffic signal (h),

r_{ms} is the red duration at the traffic signal (h),

C is the total capacity across all lanes at the main signal (veh/h),

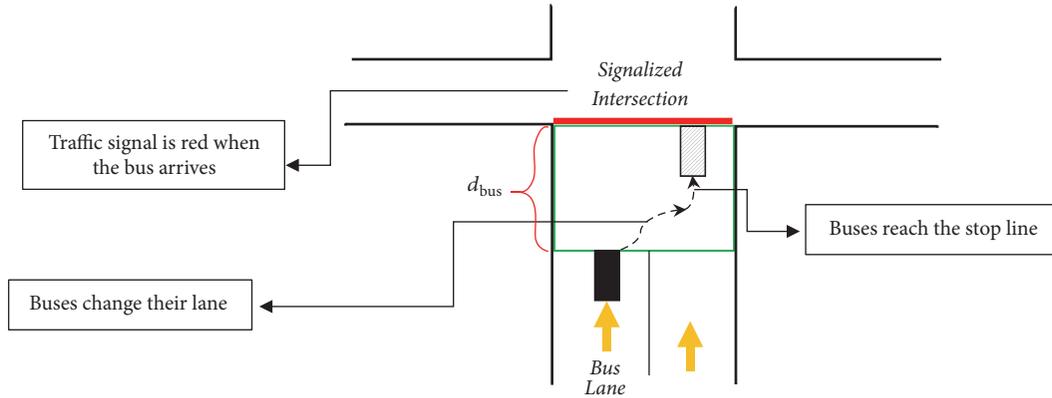


FIGURE 1: The buses' approach for changing lane and reaching the stop line.

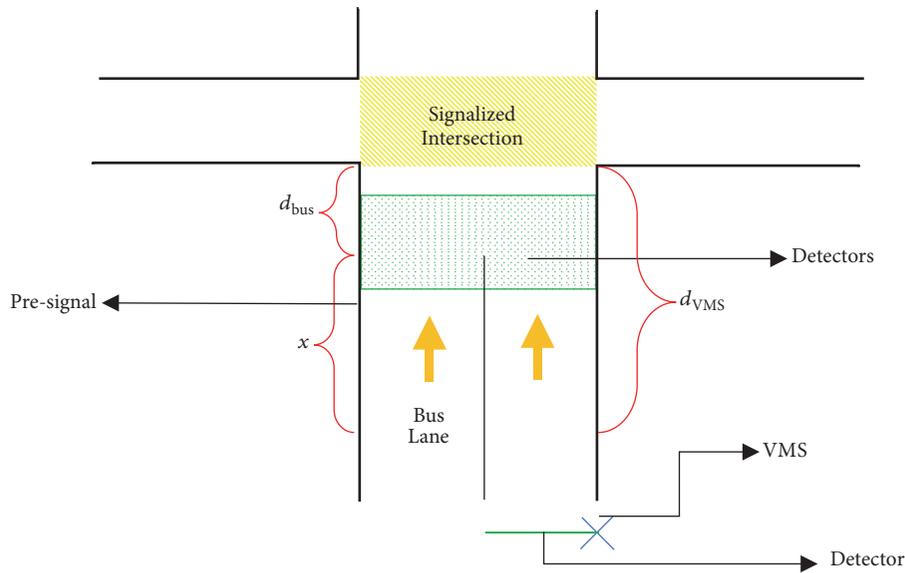


FIGURE 2: The schematic intersection components.

k_{jam} is jam density (density at zero speed) (veh/km),
 C/k_{jam} is the bus moving speed for changing lane and reaching the intersection,
 v_1 is the bus initial speed, which is not zero,
 $(c - r_{ms})$ is the green duration of the traffic signal,
 $(c - r_{ms})_1$ is the required time for buses to change the lane with initial speed of v_1 ,
 $(c - r_{ms})_2$ is the required time for buses to reach the stop line with speed of C/k_{jam} .

Step 2 (determine the distance between cars and intersection for installing a VMS in order to give the car real-time information about its appropriate speed). The required distance lets the lead vehicle accelerate or decelerate its speed. In other words, it would be an essential distance for a car to moderate and justify its speed in the critical traffic situation. This distance (4) is related to the distance between bus and

intersection, and the drivers' reaction time and their current speed. It should be considered that the distance of presignal (d_{bus}) estimated in (3) has to be added to this distance. Its details are depicted in Figure 2.

To calculate the distance:

$$x = \frac{v_{current} t_{reaction}}{3.6} \tag{4}$$

$$d_{VMS} = x + d_{bus}, \tag{5}$$

where

$v_{current}$ is the speed of the lead vehicle, which is assumed constant and based on the road rules (km/h),
 $t_{reaction}$ is the reaction time of the lead vehicle once deciding to change its speed (s), which is assumed 2 s,
 x is the distance between the lead vehicle and the presignal,

d_{VMS} is the essential distance between the lead vehicle and the intersection which is required for the lead vehicle in the critical situation to stop the car (m).

Step 3 (optimize the speed of vehicles before reaching presignals for the sake of reducing their number of stops). In order to minimize the cars' number of stops, speed and initial queue delay are considered. Alleviating the private vehicles' stops is done through controlling the vehicles speed, which leads to eliminating the vehicles' initial queue delay. When the bus arrives at d_{bus} , cars would be declared about its lane changing. As mentioned above, cars get the alarm earlier in d_{VMS} . Thus, cars get the chance to reduce their speed pertaining to several factors: buses travel time, speed, selected lane, d_{bus} , traffic signal timing, whether it is red or green, and the performance of the buses while crossing the intersection. Using the mentioned factors, the desired arrival time for cars is estimated. Then its speed can be calculated in accordance with the distance, d_{VMS} .

In order to estimate the vehicle's speed and consequently the following vehicles' speed, the below equations are used:

$$t_B = \max [t_{bus}, r_{ms}]$$

$$t_{bus} = \frac{3.6d_{bus}}{v_{bus}}, \quad (6)$$

where

t_B is the bus's travel time according to traffic signal timing (s),

t_{bus} is the bus's travel time (for changing its lane) to pass the area behind the stop line (s),

d_{bus} is the distance between the presignal and the main intersection (m),

v_{bus} is the bus's mean speed of near the intersection (km/h),

r_{ms} is the duration of red time in traffic signal (s).

$$t_{car} = \frac{3.6d_{VMS}}{v_{car,current}}, \quad (7)$$

t_{car} is the car's travel time to reach the intersection (s),

d_{VMS} is the proposed distance for installing the VMS (m),

$v_{car,current}$ is the car's mean speed in arterials (km/h).

$$t_C = t_B + t_{car}, \quad (8)$$

t_C is the car's travel time according to the leading traffic condition (s).

$$v_{car,suggested} = \frac{d_{VMS}}{t_C} \times 3.6 = v_{lead}, \quad (9)$$

$v_{car,suggested}$ is the suggested speed of the car in arterials according to the leading traffic condition, in terms of the existence of the buses changing their lanes and traffic signal timing (km/h).

The suggested speed for private vehicles estimated in (9), $v_{car,suggested}$, is the lead vehicle's speed. As a result, the private vehicles are informed of the traffic condition prior to reaching the intersection, and they are notified about the appropriate speed. Consequently, not only the number of stops for private vehicles would be reduced, but also the buses would be prioritized to select their approach.

The point is that the VMS should be installed in the optimum place to inform the private vehicles in an appropriate distance. This distance should be efficient in informing private vehicles even in the critical traffic condition. It should be considered that traffic condition is influenced by several factors which are as follows:

- (i) Occupying the mentioned distance by private vehicles travel in front of the lead vehicle.
- (ii) The bus, which travels in the bus lane, changes its lane in d_{bus} distance, or keeps moving in its lane.
- (iii) Considering time and type of traffic signal in terms of red or green and its duration.

The value of speed depicted in the VMS needs to be updated when the traffic signal changes from red to green or vice versa.

4. Results and Discussion

The proposed model would be analyzed numerically through an example of signalized intersection equipped with a presignal system for both before and after implementing the model. In addition, vehicles' delay could be estimated in the mentioned cases, and the desirability of the model would be demonstrated by comparing the results. For the sake of analyzing this model, some assumptions have been assumed which are presented in Table 1. More specifically, the signal cycle length is 60 s, and the ratio of green time to cycle length (g/c) is 0.5. The density and capacity of the arterial are 100 veh/km and 800 veh/h, respectively. In addition, the proportion of traffic flow to capacity (V/C) is 1. The speed of buses and cars is hypothetically 20 km/h and 40 km/h before implementing the model.

The essential distance for buses to change their lane before the intersection (d_{bus}) and the required distance between the lead vehicle and the intersection for installing VMS (d_{VMS}) could be calculated using the mentioned parameters (assumed in Table 1) and the equations (proposed in the previous section). On the other hand, suggested speed and travel time of the cars would be estimated to be shown on VMS.

The most important difference between the current situation and applying the proposed model is forming a queue of vehicles behind presignals due to the red signal imposed on cars. This issue would cause initial queue delay in the case of implementing presignal without any modification. Consequently, there would be a remarkable increase in delay. In other words, implementing the model would eliminate the cars' stops behind the presignal. Therefore, the cars' delay and per person's delay related to additional stop would be alleviated. As it has been shown in Table 1, the delay has been reduced 20% through implementing the model in $V/C = 1$

TABLE 1: Assumptions for parameters used in the model.

Parameters	Values (current situation)	Values (after implementing the model)
Duration	1 h	1 h
Number of buses	12	12
Number of cars (V)	800 veh	800 veh
Bus headway	300 s	300 s
Car headway	4.5 s	4.5 s
Cycle length (c)	60 s	60 s
Red duration (r_{ms})	30 s	30 s
Capacity (C)	800 veh/h	800 veh/h
Jam density (K_{jam})	100 veh/km	100 veh/km
Initial speed of bus (V_{1b})	20 km/h	20 km/h
Initial speed of car (V_{1c})	40 km/h	40 km/h
Green duration (G)	30 s	30 s
Time for bus to change lane	15 s	15 s
Time for bus to reach the stop line	15 s	15 s
Capacity/density	8	8
d_{bus}	83.33 m	83.33 m
Distance of VMS for car (x)	--	22.22 m
Reaction time (t)	2 s	2 s
d_{VMS}	--	106 m
Bus travel time (t_{bus})	15 s	15 s
Suggested travel time of bus (t_B)	30 s	30 s
Car travel time (t_{car})	--	9.5 s
Suggested travel time of car (t_C)	--	39.5 s
Suggested speed of car ($V_{car, suggested}$)	--	9.62 km/h
Traffic flow (V)	800 veh/h	800 veh/h
$X (V/C)$	1.00	1.00
Initial queue ($Q_{(b)}$)	10 veh	0 veh
Uniform delay ($d_{(1)}$)	15 s	15 s
Incremental delay ($d_{(2)}$)	64 s	64 s
Progression adjustment factor (PF)	1.667	1.667
Initial queue delay ($d_{(3)}$)	22.5 s	0 s
Delay	111.14 s	88.64 s
Travel time bus	126.14 s	103.64 s
Travel time car	150.64 s	128.14 s
g/c	0.5	0.5
Average car occupancy	2 p	2 p
Delay/person	55.57 s/pp	44.32 s/pp

and $g/c = 0.5$, and it would be an illustration for the efficient effect of the model on both cars' delay and per person's delay.

In order to conduct numerical analyses of the model, the fluctuation of g/c and V/C would be considered, and its influence on cars' delay would be determined. Figure 3 compares the amount of delay in the current situation and after implementing the model at $g/c = 0.5$ and different V/C ratios.

As is shown in Figure 3, implementing the proposed model has declined delay at all V/C ratios. It has to be mentioned that the percentage of delay improvement is decreased as the ratio of V/C increased. In other words, the reduction of delay is 63% at undersaturated condition, while it is 20% at oversaturated one.

5. Conclusion

One of the negatively influential issues regarding urban life is the traffic jam. It has negative impacts on people's life and well-being due to increasing their travel time and delay. It has to be mentioned that it would not only disturb the dwellers' planning and timetable, but also threaten individuals' health because of air pollution.

The suggested strategy to decrease traffic jam is to encourage people to use public transportation through improving its systems. One of these motivations is to give priority to this mode by implementing presignals behind the signalized intersections. Presignal is a signal that is installed in advance of the intersections and gives private cars red signal to

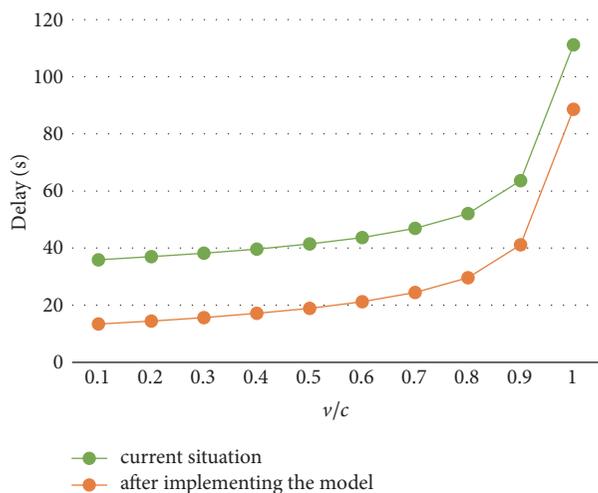


FIGURE 3: Comparison of private vehicles' delay for the current situation and the proposed presignal's model with different V/C ratios at $g/c = 0.5$.

prioritize buses in selecting their lane and approach behind the main intersection. By this way, conflicts between cars and buses while discharging the intersection are eliminated and the priority for locating in the first line of the queue and discharging the intersection is given to buses. It should be considered that as buses have much higher occupancies, the overall delay and travel time for passengers are decreased remarkably.

This strategy reduces buses' delay and travel time and increases their speed efficiently. However, cars are not taken into account in presignals, and they are given additional stops behind the presignal. More specifically, installing presignals enhance the functionality of buses by giving them priority and eliminating, or at least alleviating, the probable conflicts between them and cars. However, it worsens the performance of the private vehicles because of increasing their number of stops.

The study aims to develop this relatively modern method, presignals. In this research, it is considered to minimize the private vehicles' number of stops behind presignals. Thus, their travel time, delay, and speed are optimized. In other words, it is suggested to prioritize buses and develop cars' performance simultaneously by minimizing their number of stops behind presignals. For this purpose, it is proposed to balance the cars' speed according to buses' destination and traffic signal timing. Therefore, the delay of cars would be decreased remarkably.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study is supported by the National Natural Science Foundation of China (no. 71701070) and the Natural Science Foundation of Guangdong Province (no. 2016A030310427).

References

- [1] L. Han, D. Z. W. Wang, H. K. Lo, C. Zhu, and X. Cai, "Discrete-time day-to-day dynamic congestion pricing scheme considering multiple equilibria," *Transportation Research Part B: Methodological*, vol. 104, pp. 1–16, 2017.
- [2] Y. Kuang, X. Qu, and S. Wang, "Propagation and dissipation of crash risk on saturated freeways," *Transportmetrica B*, vol. 2, no. 3, pp. 203–214, 2014.
- [3] D. Z. W. Wang and B. Du, "Continuum modelling of spatial and dynamic equilibrium in a travel corridor with heterogeneous commuters—A partial differential complementarity system approach," *Transportation Research Part B: Methodological*, vol. 85, pp. 1–18, 2016.
- [4] Z. Liu, S. Wang, B. Zhou, and Q. Cheng, "Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 58–72, 2017.
- [5] X. Qu, S. Wang, and J. Zhang, "On the fundamental diagram for freeway traffic: a novel calibration approach for single-regime models," *Transportation Research Part B: Methodological*, vol. 73, pp. 91–102, 2015.
- [6] Z. Liu, Y. Yan, X. Qu, and Y. Zhang, "Bus stop-skipping scheme with random travel time," *Transportation Research Part C: Emerging Technologies*, vol. 35, pp. 46–56, 2013.
- [7] A. Nayan and D. Z. W. Wang, "Optimal bus transit route packaging in a privatized contracting regime," *Transportation Research Part A: Policy and Practice*, vol. 97, pp. 146–157, 2017.
- [8] Y. Bie, D. Wang, and X. Qu, "Modelling correlation degree between two adjacent signalised intersections for dynamic subarea partition," *IET Intelligent Transport Systems*, vol. 7, no. 1, pp. 28–35, 2013.
- [9] Z. Liu, S. Wang, W. Chen, and Y. Zheng, "Willingness to board: a novel concept for modeling queuing up passengers," *Transportation Research Part B: Methodological*, vol. 90, pp. 70–82, 2016.
- [10] Q. Meng and X. Qu, "Bus dwell time estimation at bus bays: a probabilistic approach," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 61–71, 2013.
- [11] Y. Bie, Z. Liu, and Y. Wang, "A real-time traffic control method for the intersection with pre-signals under the phase swap sorting strategy," *PLoS ONE*, vol. 12, no. 5, Article ID 0177637, 2017.
- [12] M. Peake, "Getting buses to the front of the queue," in *Proceedings of the IPENZ Transportation Conference*, Queenstown, New Zealand, 2006.
- [13] J. Wu and N. Hounsell, "Bus priority using pre-signals," *Transportation Research Part A: Policy and Practice*, vol. 32A, no. 8, pp. 563–583, 1998.
- [14] S. S. P. Kumara and N. B. Hounsell, "Bus priority and queue relocation capabilities of pre-signals: an analytical evaluation," in *Proceedings of the 11th IFAC Symposium on Control in Transportation Systems*, pp. 391–396, Delft, The Netherlands, 2006.
- [15] H. He, S. I. Guler, and M. Menendez, "Adaptive control algorithm to provide bus priority with a pre-signal," *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 28–44, 2016.
- [16] Y. Xuan, *Increasing the flow capacity of signalized intersections with pre-signals: theory and case study [PhD thesis]*, University of California Berkeley, 2011.
- [17] Y. Xuan, M. J. Cassidy, and C. F. Daganzo, "Using a pre-signal to increase bus- and car- carrying capacity at intersections: theory and experiment," in *TRB 2012 Annual Meeting*, 2011.

- [18] L. Kejun, "Bus priority signal control at isolated intersection," in *International Conference on Intelligent Computation Technology and Automation*, vol. 2, pp. 234–237, IEEE Computer Society, Montpellier, France, 2008.
- [19] H. He, S. I. Guler, and M. Menendez, "Providing bus priority using adaptive re-signals," *Online Pre-Signal Strategies TRB - Final after Reviews*, 2014, <http://docs.trb.org/prp/15-2793.pdf>.
- [20] S. I. Guler and M. Menendez, "Analytical formulation and empirical evaluation of pre-signals for bus priority," *Transportation Research Part B: Methodological*, vol. 64, pp. 41–53, 2014.
- [21] S. I. Guler and M. Menendez, "Pre-signals for bus priority: basic guidelines for implementation," *Public Transport*, vol. 7, no. 3, pp. 339–354, 2015.
- [22] S. I. Guler, V. V. Gayash, and M. Menendez, "Bus priority at signalized intersections with single-lane approaches: A novel pre-signal strategy," *Transportation Research Part C: Emerging Technologies*, vol. 63, pp. 51–70, 2016.
- [23] *Highway Capacity Manual: Metric Units*, Transportation Research Board, Washington, DC, USA, 2000.

Research Article

An Association Rule Based Method to Integrate Metro-Public Bicycle Smart Card Data for Trip Chain Analysis

De Zhao ^{1,2}, Wei Wang,³ Ghim Ping Ong,⁴ and Yanjie Ji ³

¹Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou No. 2, Nanjing 210096, China

²Department of Civil and Environmental Engineering, National University of Singapore, Engineering Drive 2, E1A 08-20, Singapore 117576

³School of Transportation, Southeast University, Si Pai Lou No. 2, Nanjing 210096, China

⁴Department of Civil and Environmental Engineering, National University of Singapore, Engineering Drive 2, E1A 07-03, Singapore 117576

Correspondence should be addressed to De Zhao; zhaode.0@aliyun.com

Received 26 January 2018; Accepted 17 April 2018; Published 28 May 2018

Academic Editor: Lele Zhang

Copyright © 2018 De 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.

Smart card data provide valuable insights and massive samples for enhancing the understanding of transfer behavior between metro and public bicycle. However, smart cards for metro and public bicycle are often issued and managed by independent companies and this results in the same commuter having different identity tags in the metro and public bicycle smart card systems. The primary objective of this study is to develop a data fusion methodology for matching metro and public bicycle smart cards for the same commuter using historical smart card data. A novel method with association rules to match the data derived from the two systems is proposed and validation was performed. The results showed that our proposed method successfully matched 573 pairs of smart cards with an accuracy of 100%. We also validated the association rules method through visualization of individual metro and public bicycle trips. Based on the matched cards, interesting findings of metro-bicycle transfer have been derived, including the spatial pattern of the public bicycle as first/last mile solution as well as the duration of a metro trip chain.

1. Introduction

Public bicycle usage for metro access provides new opportunities for sustainable transportation, helping to address the “first-mile” and “last-mile” problems [1]. To understand the effect of integrating public bike and metro systems, transportation planners and researchers have been striving to evaluate the transfer efficiency and behavior through personal travel profiling [2], social-demographic information [1, 3], or public bicycle historical trips [4]. Previous attempts to understand metro and public bicycle transfer are limited, partly due to the difficulty in data collection. Conventional household travel surveys or diaries are time-consuming and laborious to carry out while convenient access to large travel datasets and integration across different data platforms are yet to be found.

In China, both metro and public bicycle transactions are made via Automatic Fare Collection (AFC) system, also

known as smart card (SC). SC data, with a massive sample size, can provide valuable insights into the understanding of metro-public bicycle transfer behavior [5, 6]. Compared with conventional surveys, SC data collection, being a by-product of revenue collection, is a convenient method for retrieving travel patterns of commuters. Therefore, a great deal of research studies have emerged in terms of SC data mining [6–12].

However, the smart cards for metro and public bicycle systems are issued and managed by independent companies and in most cases, commuters need to hold two different smart cards to complete a public bicycle-metro trip chain. As such, the metro SC dataset and public bicycle SC dataset are saved independently without a common and unique identifier (ID) for a single commuter. This makes it difficult for researchers and transit agencies to leverage big SC data to investigate metro-bike transfer behavior effectively and efficiently, unless there exists a method to match the unique

card IDs within each system. It is not possible to directly and accurately match the smart cards from different datasets, as accessing personal information may offend the privacy. Nevertheless, the detailed individual travel pattern hidden in the SC data makes it possible to match card IDs of the same commuter.

Therefore, the primary objective of this study is to develop a data fusion methodology for matching metro and public bicycle smart cards of commuter identity in an integrated metro-public bicycle network. To achieve the aim, this study provides a novel approach using association rules (AR), a concept in the machine learning domain. The smart card data from Nanjing metro and public bicycle in China is used to demonstrate and validate our developed method. The remainder of the paper is structured as follows. In the literature review section, previous studies on smart card data for trip chain, multisource data fusion, and association rules are reviewed. The methodology and data source section formulates the association rules to match smart card travel and identifier data and presents the data source preprocessing procedures and validation approach. The results and validation section applies the proposed data fusion method to empirical analysis, calibrates the key parameters, and validates the proposed method. The section also presents possible application after matching metro and public bicycle card IDs. The conclusions and recommendations section concludes the paper and gives the research limitations as well as recommendations for future study.

2. Literature Review

Transit smart card data records transit riders' detailed trip log, which can be used to analyze the transit riders' trip chain. A huge body of literature has grown with regard to SC data analysis [7, 11, 14–17]. Means of mining SC data are various, e.g., data fusion and machine learning. Three streams of research are relevant to this study: (1) smart card data for trip chain; (2) multisource data fusion; (3) association rules.

2.1. Smart Card Data for Trip Chain. Past research studies in the literature have analyzed historical SC data to estimate transit origin location [18], destination location [6, 19], and total daily or monthly transit trip chain pattern [8, 10]. Furthermore, long-term year-to-year changes in transit users' trip habits could also be tracked and analyzed [12]. As for public bicycle, the research just started in recent years [20]. Notably, most of the research used bicycle trip data [21] rather than true SC data, since the trip data is easier to obtain. Trip data is usually open to the public in cities of United States or Europe, where the bicycle rental is accomplished via credit card or cell phone app. In general, compared with SC data, bicycle trip data lacks card ID and thus cannot be used to model users' travel behavior. By far, public bicycle SC data has been used to investigate public bicycle users' travel patterns [22] as well as bicycle trip chains for men and women [14, 23]. Public bicycle SC data could also help to classify different types of behaviors and compare the trip disparity [24].

2.2. Multisource Data Fusion. SC data provides much detailed information about each trip, but not the information about trip purpose, user assessment, and ultimate destination. When integrated with other data sources, SC data can play a greater role in mining transit riders' behavior and validating previous research approaches. By integrating both SC and Global Positioning System (GPS) data, Munizaga and Palma [9] estimated the OD of multimodal transit systems and validated the results against metro OD surveys in Santiago, Chile [16]. Ma and Wang [25] built a data-driven platform by integrating SC and GPS data to monitor transit performance in Beijing. Researchers can also examine the spatial-temporal dynamics of bus passengers and estimate the trip purposes when matching SC data, respectively, with General Transit Feed Specification (GTFS) data [15] and person trip survey data [17]. Yet, very little research attempted to integrate metro SC data with public bicycle SC data for investigating metro-bicycle transfer.

2.3. Association Rules. AR was first introduced by Agrawal et al. [26], and they applied this model to the supermarket transaction data to find out what items people would buy together. They also proposed algorithms for finding the AR. Shortly after that, the method was applied to other fields as a popular machine learning technique, including transportation. AR was firstly used in the transportation area by Keuleers et al. [27] to learn the travel patterns of multiday activity diaries. Soon after that, Keuleers et al. [28] tried to recognize temporal effects that may exist in the same data. AR showed high efficiency and convenience in rules mining. Later, Kusumastuti et al. [29] explored individuals' thoughts about leisure-shopping travel decisions by means of AR. Diana [30] used AR analysis to explore travel patterns of different modes based on 2009 US National Household Travel Survey and found the substitution effect between private modes and public transit. In particular, Chu and Chapleau [31] used AR to mine behavior rules of SC users and found some potential regularities with a high level of confidence.

Among all the relevant studies presented in this section, there still remains lack of a data fusion methodology to match smart cards from different sources. As AR is capable of identifying potential relationships between items, this paper attempts to develop AR-based algorithm to match metro and public bicycle SC data of the same person within an integrated bus-public bicycle network. We convert metro SC data and public bicycle SC data into transaction datasets and follow the method of Agrawal et al. [26] to match the card IDs. In our paper, we also propose an approach to validate the developed method.

3. Methodology and Data Source

3.1. Association Rules. Let $I = \{i_1, i_2, \dots, i_k\}$ be a set of items. A transaction d_i is defined as a group of items, namely, a subset of I . $D = \{d_1, d_2, \dots, d_n\}$ is a set of all transactions called the transaction database. Each transaction d_i in D has a unique transaction number. An association rule is used to describe potential relations of several items in the transaction

```

(1)  $L_1 = \{\text{large 1 - itemsets}\};$ 
(2) for  $(k = 2; L_{k-1} \neq \phi; k++)$  do begin
(3)    $C_k = \text{apriori-gen}(L_{k-1});$  //New candidates
(4)   for all transactions  $t \in D$  do begin
(5)      $C_t = \text{subset}(C_k, t);$  //Candidates contained in  $t$ 
(6)     for all candidates  $c \in C_t$  do
(7)        $c.\text{count}++;$ 
(8)   end
(9)    $L_k = \{c \in C_k | c.\text{count} \geq \text{min support}\}$ 
(10) end
(11) All frequent sets  $= \bigcup_k L_k$ 

```

ALGORITHM 1: Apriori algorithm [13].

database D and is expressed as $X \Rightarrow Y$, where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. X is called antecedent or left-hand side (LHS), and Y is called consequent or right-hand side (RHS). For example, in supermarket sales data mining, the rule {butter, bread} \Rightarrow {milk} means if a customer buys both “butter” and “bread”, he is also likely to buy “milk”. In this research, we set all metro card IDs and public bicycle card IDs as items I . An association rule would indicate that there is potential association between card IDs in LHS and RHS. To better match two smart cards by AR, we should try our best to cluster two cards of the same person into one transaction d_i .

In association rules, there are three key parameters: support, confidence, and lift. The corresponding definitions are listed below. The support value of X , represented as $\text{supp}(X)$, means the probability that the item-set X appears in the database D , defined as the proportion of transactions that includes the item-set X in D , as in (1). Accordingly, $\text{supp}(X \Rightarrow Y)$ can be expressed by (2). The generalized expression $X \cup Y$ in association rules means the union of the items in X and Y rather than either X or Y :

$$\text{supp}(X) = P(X) = \frac{\text{num}(X)}{\text{num}(D)}, \quad (1)$$

$$\text{supp}(X \Rightarrow Y) = P(X \cup Y) = \frac{\text{num}(X \cup Y)}{\text{num}(D)}. \quad (2)$$

The confidence value of a rule, expressed as $\text{conf}(X \Rightarrow Y)$, indicates the proportion of the transactions containing both X and Y in those that contain X :

$$\text{conf}(X \Rightarrow Y) = P(Y | X) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}. \quad (3)$$

The lift value is used to describe the effectiveness of the association rule, as defined in (4). A lift of 1.0 implies that the occurrence of X has nothing to do with that of Y . That is to say, no association rule can be found between X and Y when the lift = 1.0. When the lift is more than 1.0, $X \Rightarrow Y$ is an effective association rule and greater value of lift indicates stronger association rule:

$$\text{lift}(X \Rightarrow Y) = \frac{P(Y | X)}{P(Y)} = \frac{\text{supp}(X \cup Y)}{\text{supp}(X) \times \text{supp}(Y)}. \quad (4)$$

The Minimum Support (MS) value and the Minimum Confidence (MC) value are set as a constraint on measure of significance to ensure that the rules under consideration are sufficiently significant. MS is used to search the most frequent item-sets and MC is used to form rules based on these frequent item-sets. The former process is computationally intensive. To accomplish the former process, the commonly used Apriori algorithm is applied, as shown in Algorithm 1. As the Apriori algorithm can be found in previous research [13, 27, 28], we do not expatiate in this study.

3.2. Data Source and Preprocessing. Two major datasets were used in this study: metro SC data and public bicycle SC data, as shown in Figures 1(a) and 1(b). The datasets were recorded from November 1, 2015, to November 24, 2015, obtained from Nanjing Smart Card Company and Public Bicycle Company, respectively. The metro SC data contains the card ID, departure station, tap-in time, arrival station, and tap-out time. The original public bicycle data contains card ID, rent station, rent time, return station, and return time. Based on the location of the bicycle rent/return station, we added its nearby metro station in the 300 m buffer to the data frame. A buffer of 300 m radius was used as the walkable distance for public bicycle trips [32], because the planning standard promulgated by Nanjing government has suggested the walkable distance for community public facility is 300 m (5-min walk) [33].

The 24-day metro SC data contains over 34 million rows of trips and the 24-day public bicycle SC data includes nearly 1.2 million rows of trips. To mine AR between two databases, we need to merge and convert the SC data into transaction data. As mentioned above, we need to cluster two cards of the same person into one transaction to the greatest extent possible. The most likely transaction is metro-bicycle transfer or bicycle-metro transfer.

Firstly, we divide metro SC data and public bicycle SC data into subsets of regular time slot (TS) based on tap-in time and return time, respectively. Then, we merged the metro SC subset and the public bicycle SC subset with the same time slot and metro station (departure station and return nearby metro station, respectively) as one transaction. For bicycle-metro transfer, two cards of the same person may probably appear in the same transaction. Similarly, we created the metro-bicycle transfer transactions by cutting and merging

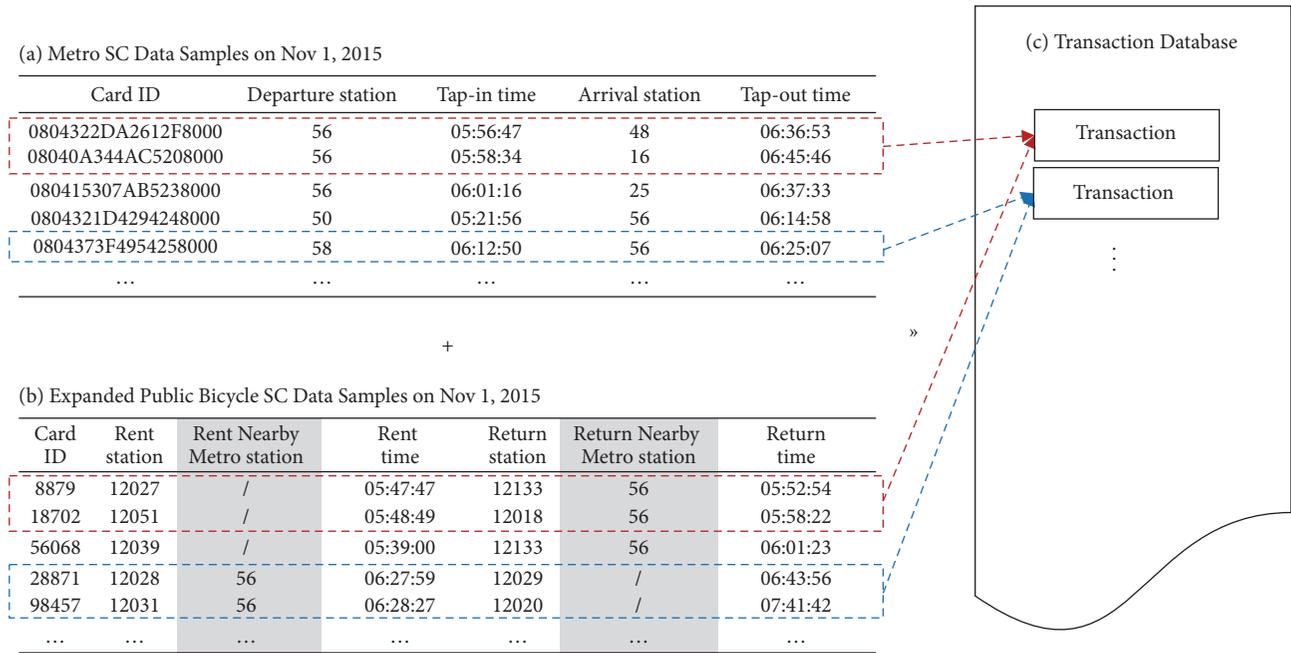


FIGURE 1: Smart card data and transaction dataset generation.

the datasets based on tap-out time, rent time, arrival station, and rent nearby metro station. All the transactions added up to create the transaction database D , as shown in Figure 1(c). In order to reduce the computational burden, we removed transactions with only metro card IDs or only public bicycle card IDs, because such transactions only contribute to finding internal relationship among metro cards or among public bicycle cards.

3.3. Validation Approach. Unfortunately, we cannot directly know whether two cards belong to the same person, because the only way to identify one person between different databases is to obtain nonanonymous personal information, which may violate individual privacy. Instead, we put forward a surrogated approach to validate the results: using the data in the first 20 days (train data) to train the association rules and the data in the last 4 days (test data) to validate the results. We assume the matched two card IDs do belong to the same person if they meet all the following three conditions.

(1) A transfer behavior of the two card IDs is also observed in the test data. A transfer behavior is defined as renting a public bicycle within 10 minutes after exiting the metro station or entering the metro station within 10 minutes after returning a public bicycle. We use 10 minutes as the maximum value of metro-bicycle (or bicycle-metro) transfer time. As mentioned above, the walkable distance between metro stations and public bicycle stations is 300 m, which is also equivalent to 5-min walk for an average person [33]. We set the maximum value of the transfer time as twice of 5 minutes.

(2) No time-overlap was observed between the matched two card IDs in 24-day datasets. In other words, one cannot

take the metro while renting the public bike or rent a public bike during the metro ride.

(3) One metro SC ID only matches with one public bicycle SC ID. We assume each person usually owns one metro SC or public bicycle SC. Therefore, it makes sense only if one item (card ID) associated with only one other item.

Meeting all above three conditions by chance is quite a small probability event. Because there are over 1.8 million unique card IDs in the metro SC database and over 0.1 million unique card IDs in the public bicycle SC database. Based on our data, we randomly choose one card ID from the metro SC database and one card ID from the public bicycle SC database to check if they can meet the three conditions at the same time. We repeated 10,000 times of the selections; there are only 16 pairs of card IDs meeting all the validation conditions. Hence, the probability of meeting all above three conditions by chance is 0.16%.

4. Results and Validation

After cutting the SC datasets into thousands of transactions by the proposed method and learning the data with AR, associations between SC IDs are retrieved. The Apriori algorithm is used to identify the most frequent item-sets (metro SC and public bicycle SC). The associated rules learning in this research was performed using R 3.4.0. Given the three model parameters and the database of Nanjing (totally 1,149,335 items and 38,014 transactions), the average calculation time of associated rules is 18 seconds with the help of Apriori algorithm on a PC with Intel i7-6700 3.4 GHz and 16 GB DDR4 RAM.

To capture the meaningful results, we only selected rules with one metro SC ID as LHS and one public bicycle SC

TABLE 1: Extracted association rules.

No.	Association rules (LHS \Rightarrow RHS)	Support	Confidence	Lift
1	{0804322DA2612F8000} \Rightarrow {00008879}	5.78×10^{-4}	0.79	853.37
2	{08046A39EA02218000} \Rightarrow {00008429}	7.36×10^{-4}	0.97	1223.44
3	{0804142E2ABA268000} \Rightarrow {00196771}	5.52×10^{-4}	0.72	1019.53
4	{080E4711130A4D3C00} \Rightarrow {00168715}	5.52×10^{-4}	0.70	917.58
5	{0804293EE2382C8000} \Rightarrow {00151192}	7.36×10^{-4}	0.93	1182.66
6	{08042E156A85368000} \Rightarrow {00163152}	8.15×10^{-4}	0.91	936.75
...

ID as RHS. Because many-to-one rules, metro-metro rules, and bicycle-bicycle rules are all invalid, metro-metro rules probably mean that two or more cardholders often take the metro together. It is also true for bicycle-bicycle rules. These rules are not concerned in this research and thus removed from the association rules list. The extracted association rules are shown in Table 1. The metro card ID is represented as a string of 18 hex digits, while the public bicycle card ID is a string of 8 decimal digits. One AR indicates one metro SC ID matched with one public bicycle SC ID. All the extracted rules have a very high “lift” value, indicating the associations between cards are significantly strong. The support value scattered between 0 and 1.2×10^{-3} , indicating the parameter range for MS calibration.

There are three parameters MS, MC, and TS in our proposed model. They jointly determined the number of ARs (the number of matched IDs) and the accuracy of results. We need to obtain the optimal combination of three parameters in order to derive more ARs as well as better accuracy. Accuracy is defined as the ratio between the number of ARs meeting all the three validation conditions and the total number of ARs.

Figure 2 shows how the three parameters influenced the results. We set TS, respectively, as 2 min, 5 min, 10 min, and 20 min, as shown in Figures 2(a), 2(b), 2(c), and 2(d). In general, the accuracies of ARs under various parameter combinations are all very high, with most of them being over 90%. However, with increase of the number of ARs, the accuracy of results decreased. In other words, we cannot achieve the optimal levels of both number and accuracy simultaneously. To reach an accuracy of 100%, our proposed approach could at most identify 573 ARs (matched 573 pairs of SC IDs), with TS = 10 min, MS = 0.00055, and MC = 0.4 as marked in Figure 2(c). A too small a value of TS (e.g., 2 min) will reduce the maximum number of ARs because the metro card and public bicycle card of the same transfer could be divided into two transactions with high probability. On the other hand, too big TS (e.g., 20 min) could greatly improve the maximum number of ARs but may introduce many invalid ARs, thus decreasing the accuracy.

As mentioned above, there are totally 573 metro cards matched with bicycle cards under the accuracy of 100%. All the 573 paired cards have satisfied the 3 validation conditions. To ensure that the matched cards do belong to one person, we derived each individuals’ trip log and visualized six of them as Figure 3 shows. The blue solid line segment indicates metro trip, while the red dashed line segment indicates public

bicycle trip. All the metro trips and public bicycle trips of 24 days are displayed in this plot. The metro-bicycle transfer behavior of an individual can be easily identified by adjacent metro and public bicycle trip.

Apparently, Individuals 1, 2, 5, and 6 have a regular trip pattern. They used public bicycle as a daily first-/last-mile connection to metro. The majority of their metro trips are connected with public bicycle trips, showing specific symmetries between morning trip chain and afternoon/evening one. Individuals 3 and 4 are probably not traditional office workers with routine commutes, but they still take public bicycle as a good way to address the first-/last-mile problem.

After matching two cards of the same person, interesting findings of metro-bicycle transfer could be obtained. Around 2/3 of the transfers between metro and public bicycle occurred in the peak hours. Figure 4 shows the spatial analysis of public bicycle trips that connected with metro trips by peak hours of the day. The plot provides us with a visual impression of public bicycle trips as a first-/last-mile connection.

We can find that there are usually several public bicycle stops within the vicinity of metro stations and the trip demand is not shared equally among them. Users prefer to rent bicycles from only one or two of these stations, which further exacerbates the burden of rebalancing. In zones far away from the metro station, public bikes even take on longer-distance connections, for example, in the northern part of the research area. The maximum straight-line distance is over 4 km.

Another interesting finding from Figure 4 is that the first-mile trips during morning peak (7:00~9:00) have the same spatial pattern with last-mile trips during evening peak (17:00~19:00). It is also true for last-mile trips during the morning peak and first-mile trips during the evening peak. This makes sense since people who ride a public bicycle from home to metro station (first mile) in the morning tend to ride one from metro station (last mile) in the evening, and people who ride a public bicycle from metro station to workplace (last mile) in the morning tend to ride one back to metro station (first mile) in the evening.

The results of duration of the metro-bicycle trip chain are presented in Figure 5, based on the 573 matched cards. Outliers have been removed since some overlong trips will overestimate the average trip duration. The outlier is defined as any data point that is over 1.5 interquartile ranges below the first quartile or above the third quartile.

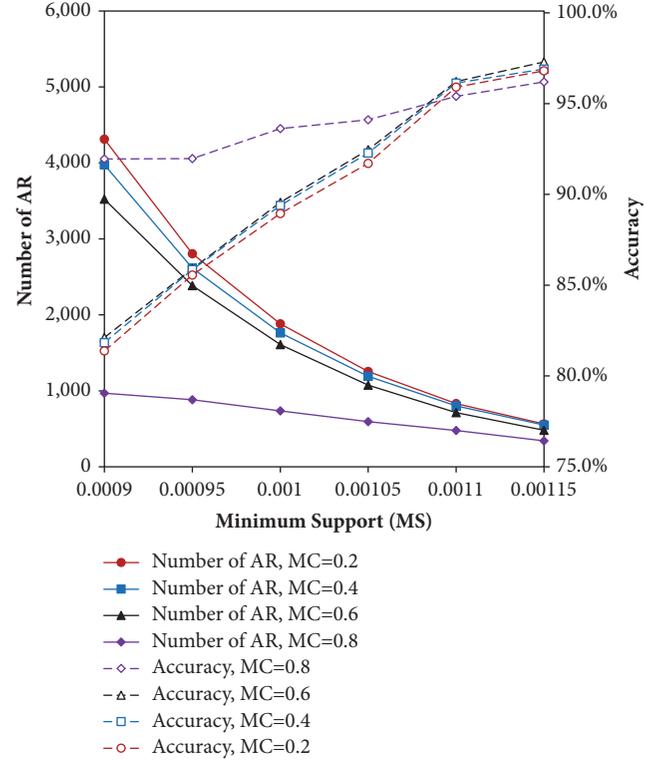
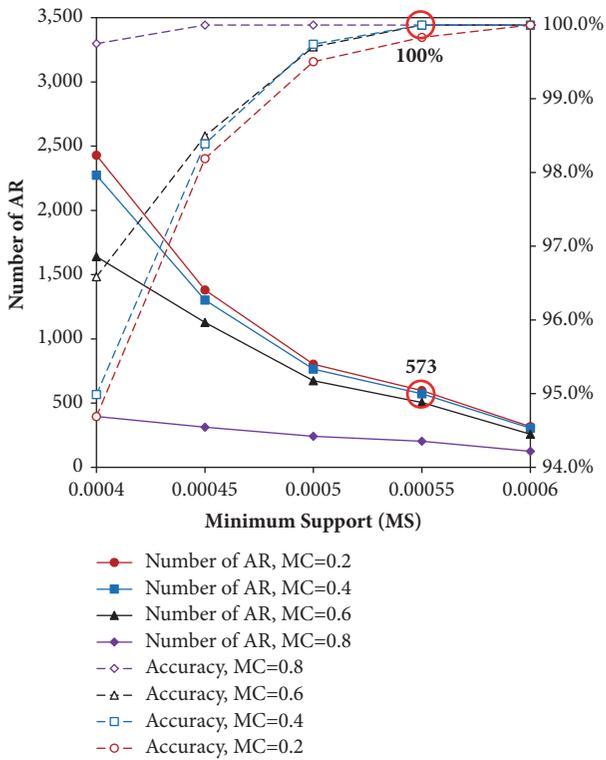
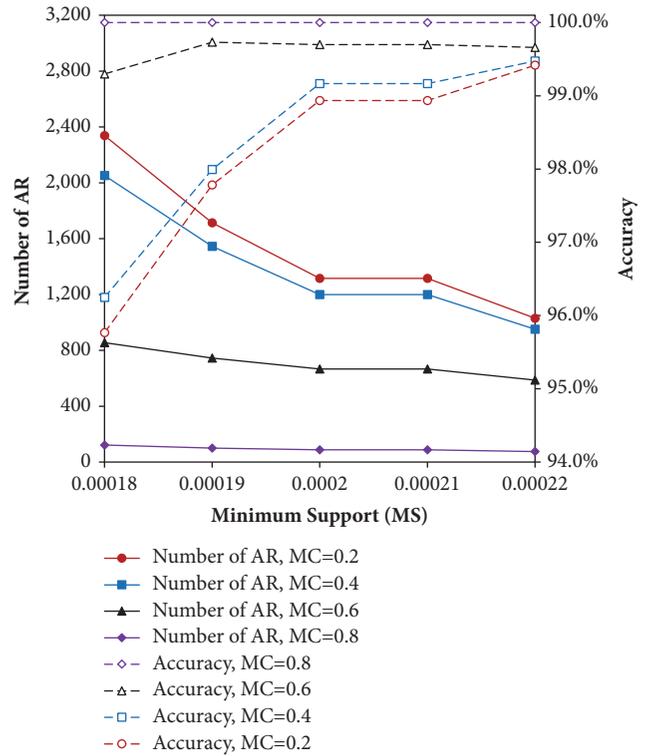
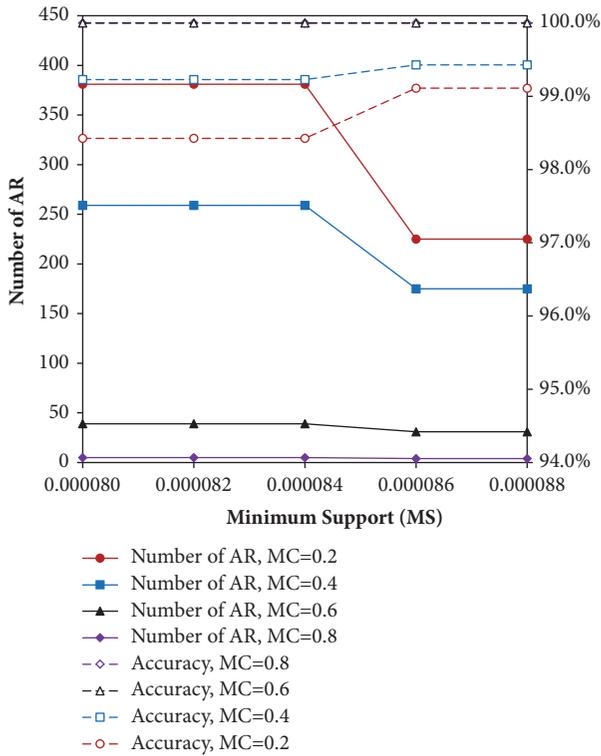


FIGURE 2: Parameter calibration.

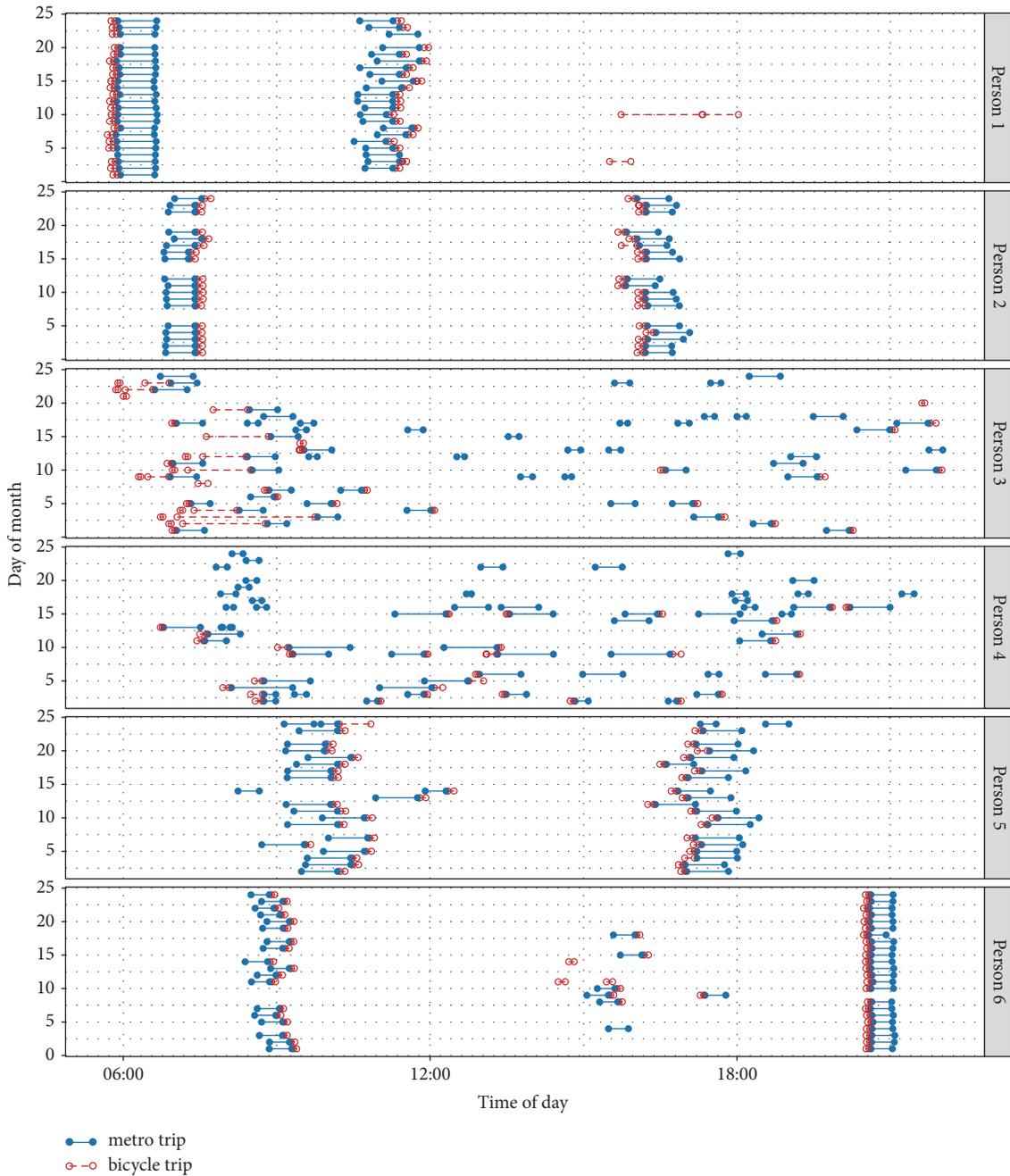


FIGURE 3: Individual trip log based on matched cards.

We divided metro-bicycle trip chain into metro trip, transfer, and public bicycle trip. For metro trip with last mile, the average metro trip time, transfer time and public bicycle time are respectively 23.59 min, 1.71 min, and 7.01 min. For metro trip with first mile, public bicycle time, transfer time and the average metro trip time are respectively 6.80 min, 1.75 min, and 23.50 min. The transfer time is very short, indicating that the walking distance between the metro station and bicycle stop is within a reasonable range. The connection time (public bicycle time and transfer time) takes around 27% of the total travel time.

When public bicycle was used as a first-mile mode, three parts of the trip chain in the evening peak are all longer than

morning peak or even nonpeak hours. This makes sense since people have to hurry to work in the morning, but they can take their time back home in the evening. Counterintuitively, for both first-mile and last-mile mode, the public bicycle trip in nonpeak hours is shorter than morning peak. This is probably because, in rush hours, people are more likely to spend much time on finding an unoccupied slot to return the bicycle.

Travel patterns comparison between matched public bicycle SC and unmatched public bicycle SC are shown in Figure 6. Outliers have also been removed to prevent overestimating the average trip duration or trip distance. The results showed that the average trip duration for matched

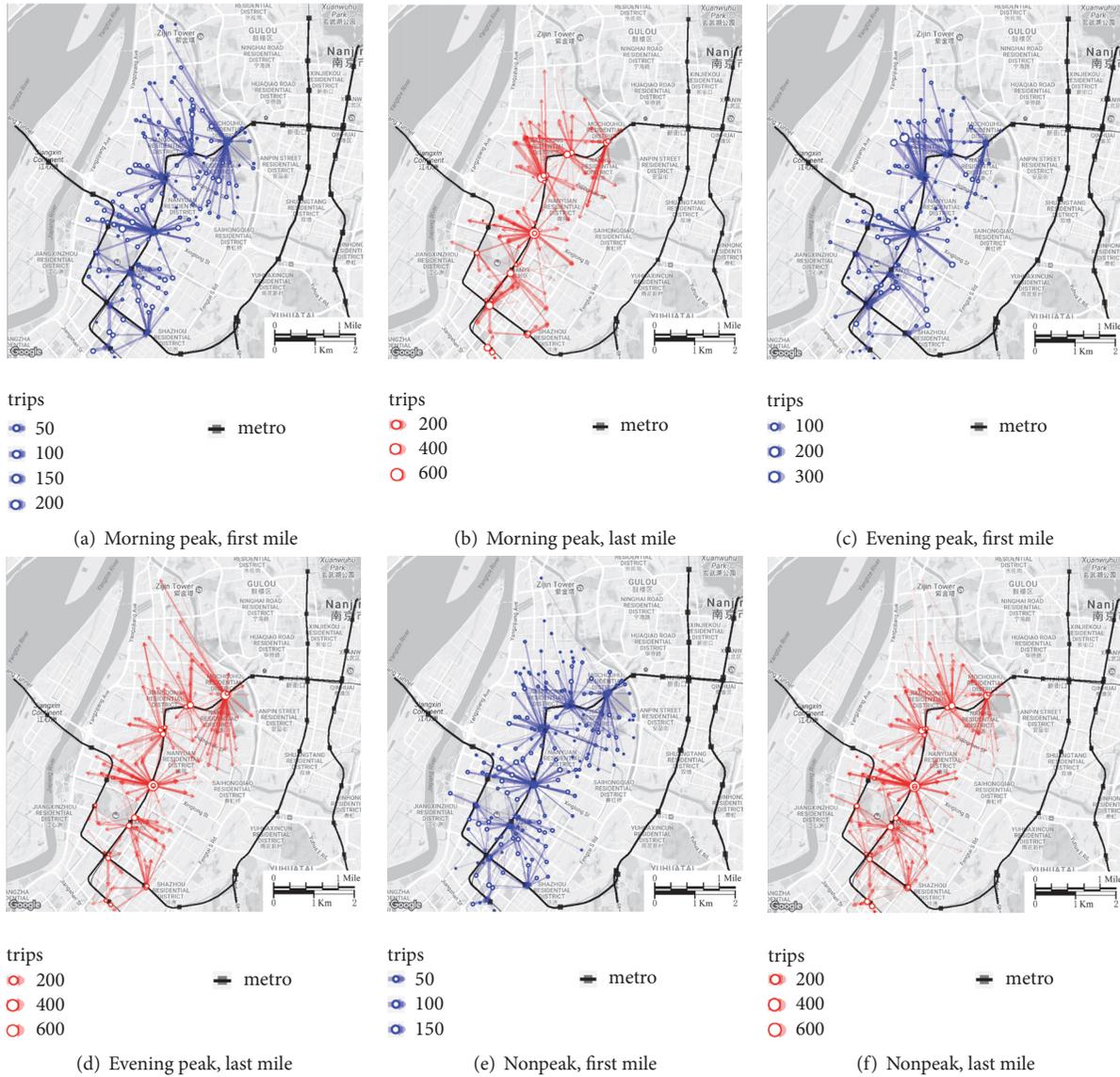


FIGURE 4: Spatial pattern of public bicycle as first-/last-mile solution.

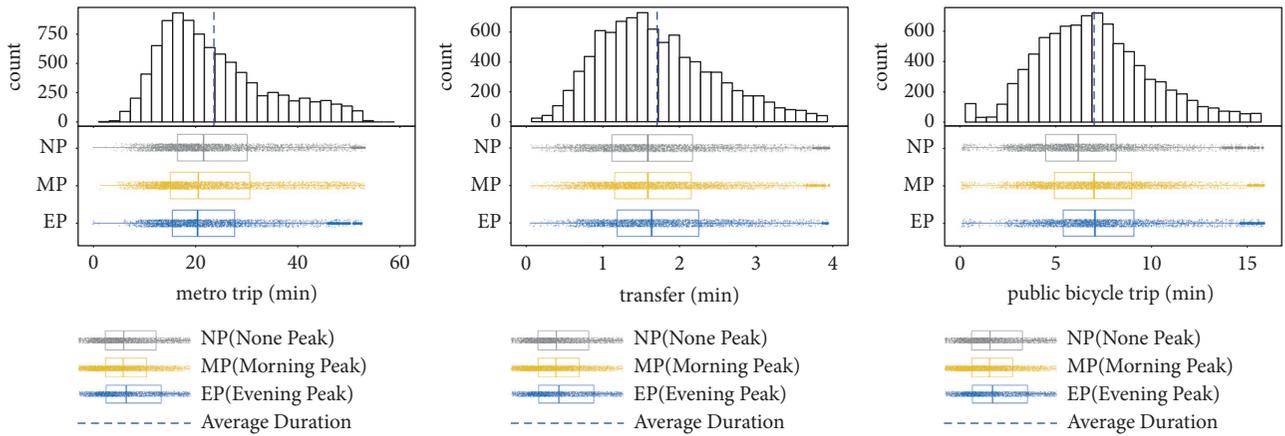
public bicycle SC data is 6.80 min, while the average trip duration for unmatched public bicycle SC data is 10.02 min with 95% of the trip duration being less than 30 min, which is consistent with previous research studies by Zhao et al. [14]. The average trip distances for matched and unmatched public bicycle SC data are 0.95 km and 1.03 km, respectively, indicating the distance of first-mile or last-mile public bicycle trips is shorter than that of the other public bicycle trips. In contrast, the average trip distance is 0.99 km in Santander, Spain [24]. Notably, unmatched public bicycle SC data also contains first-mile or last-mile trips to connect metro.

5. Conclusions and Recommendations

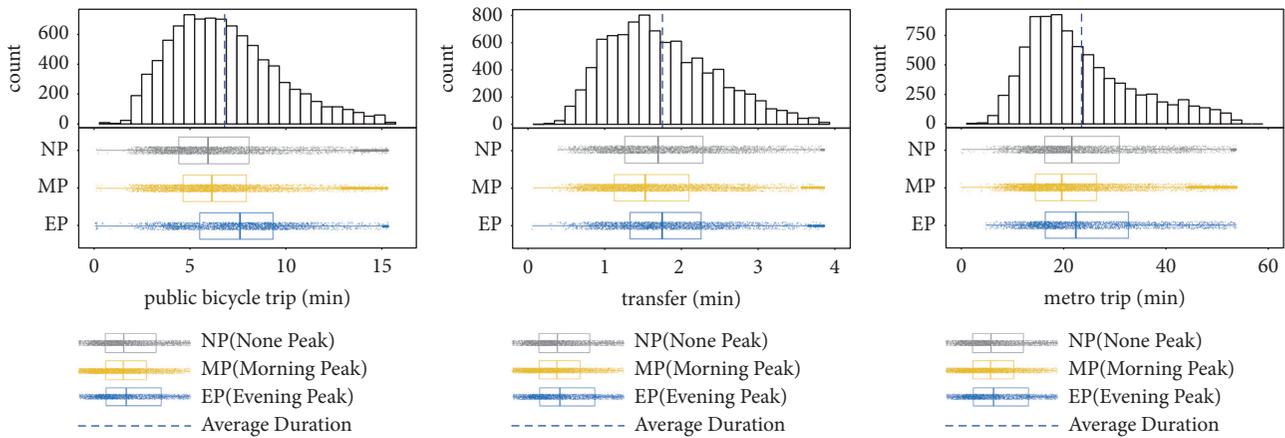
This research has put forward a novel data fusion method using association rules mining to match metro and public bicycle smart cards of the same commuter. We attempt

to match SC IDs from different sources and propose a validation approach. We calibrated the three key parameters MS, MC, and TS by demonstrating how they influenced the number of ARs and the accuracy of results. The validation process showed that, with increase of the number of ARs, the accuracy of results decreased. The individual trip log has also been derived to validate the association rules method by visualizing individual metro and public bicycle trip of each day. Based on the matched cards of the same person, interesting findings of metro-bicycle transfer have been found, including spatial pattern of public bicycle as first-/last-mile solution as well as duration of metro trip chain.

Our paper contributes to the state of knowledge by taking advantage of linked-SC data to analyze metro-bicycle transfers. We demonstrated that it is possible to match two cards of the same person based on historical SC data. Our proposed method successfully matched 573 pairs of smart cards with an accuracy of 100%, when setting three key



(a) Metro trip chain with last mile



(b) Metro trip chain with first mile

FIGURE 5: Duration of metro trip chain.

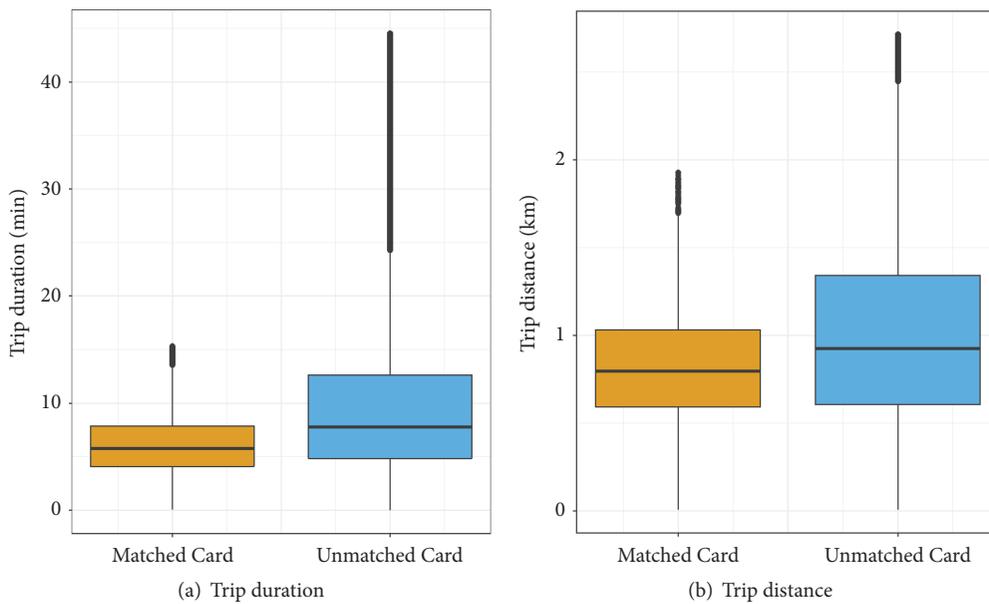


FIGURE 6: Travel patterns comparison between matched public bicycle SC and unmatched public bicycle SC.

parameters as $TS = 10$ min, $MS = 0.00055$, and $MC = 0.4$. TS determined the quality of conversion from SC data to transaction data, because TS is highly related to the metro-bicycle transfer time. However, even when transfer time is shorter than time slot (TS), the start time and end time of a transfer may still fall in two different transactions. Therefore, properly increasing the TS will help to reduce the rate of wrong grouping. Although the average transfer time is less than 1.8 min, the optimal TS of the proposed method is around 10 min.

Findings from this study suggest that around 2/3 of the transfers between metro and public bicycle in Nanjing occurred during peak hours. Although there are usually several public bicycle stops around metro station, users prefer to rent bicycles from only one or two of them, which of course exacerbates the burden of rebalancing. This paper also sheds light on the duration of the entire metro-bicycle trip chain. The connection time (public bicycle trip time and transfer time) takes around 27% of the total metro trip chain in Nanjing. When public bicycle was used as a first-mile mode, the trip duration in the evening peak is longer than morning peak or even nonpeak hours. This finding is consistent with common sense that people have to hurry to work in the morning, but they can take their time back home or dinner in the evening, while for both first-mile and last-mile mode, the public bicycle trip in nonpeak hours is even shorter than morning peak. This is probably because of the difficulty to find an unoccupied slot to return the bicycle during rush hours.

The proposed approach could be applied to other cities, where people also use different smart cards for different transportation modes. However, the parameters (MS , MC , and TS) in the model should be recalibrated based on actual SC data of the city. Because the city scale, transportation modes, and travel behaviors in different cities may greatly influence the standard of the parameters. There are several limitations of this study. Firstly, the total number of correctly matched cards were limited by the historical data. The dataset used in this research is not huge (only 24 days) due to the difficulty of obtaining both metro SC data and public bicycle SC data within the same period. Secondly, the proposed validation conditions are somewhat strict. This is because at least one metro-bicycle transfer (or bicycle-metro transfer) should occur in the remaining 4 days. Only frequent users of metro-bicycle transfer are likely to be successfully identified. However, this is the best way we can think of to validate the results without offense of individual privacy. More insightful findings about travel behaviors are expected to be found by harmonizing smart card technology across different transport modes. Speeding up the process of data integration and combining different smart cards into one will not only facilitate the traveler, but also provide data support for efficient transportation decision making.

Data Availability

The data that support the findings of this study are available from Nanjing Smart Card Company and Public Bicycle

Company but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Nanjing Smart Card Company and Public Bicycle Company.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (71701047 and 51478112) and Projects of International Cooperation and Exchange of the National Natural Science Foundation of China (5151101143).

References

- [1] Y. Ji, Y. Fan, A. Ermagun, X. Cao, W. Wang, and K. Das, "Public bicycle as a feeder mode to rail transit in China: The role of gender, age, income, trip purpose, and bicycle theft experience," *International Journal of Sustainable Transportation*, vol. 11, no. 4, pp. 308–317, 2017.
- [2] E. W. Martin and S. A. Shaheen, "Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities," *Journal of Transport Geography*, vol. 41, pp. 315–324, 2014.
- [3] J. Bachand-Marleau, J. Larsen, and A. M. El-Geneidy, "Much-anticipated marriage of cycling and transit: how will it work?" *Transportation Research Record*, vol. 2247, pp. 109–117, 2011.
- [4] R. Nair, E. Miller-Hooks, R. C. Hampshire, and A. Bušić, "Large-scale vehicle sharing systems: analysis of Vélolib," *International Journal of Sustainable Transportation*, vol. 7, no. 1, pp. 85–106, 2012.
- [5] Q. Zou, X. Yao, P. Zhao, H. Wei, and H. Ren, "Detecting home location and trip purposes for cardholders by mining smart card transaction data in Beijing subway," *Transportation*, vol. 45, no. 3, pp. 919–944, 2018.
- [6] M. Trépanier, N. Tranchant, and R. Chapleau, "Individual trip destination estimation in a transit smart card automated fare collection system," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 11, no. 1, pp. 1–14, 2007.
- [7] M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: a literature review," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 557–568, 2011.
- [8] C. Morency, M. Trépanier, and B. Agard, "Measuring transit use variability with smart-card data," *Transport Policy*, vol. 14, no. 3, pp. 193–203, 2007.
- [9] M. A. Munizaga and C. Palma, "Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile," *Transportation Research Part C: Emerging Technologies*, vol. 24, pp. 9–18, 2012.
- [10] X. Ma, Y. J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1–12, 2013.
- [11] X. Ma, C. Liu, H. Wen, Y. Wang, and Y. Wu, "Understanding commuting patterns using transit smart card data," *Journal of Transport Geography*, vol. 58, pp. 135–145, 2017.

- [12] A.-S. Briand, E. Côme, M. Trépanier, and L. Oukhellou, "Analyzing year-to-year changes in public transport passenger behaviour using smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 274–289, 2017.
- [13] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*, Santiago, Chile, 1994.
- [14] J. Zhao, J. Wang, and W. Deng, "Exploring bikesharing travel time and trip chain by gender and day of the week," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 251–264, 2015.
- [15] S. Tao, D. Rohde, and J. Corcoran, "Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap," *Journal of Transport Geography*, vol. 41, pp. 21–36, 2014.
- [16] M. Munizaga, F. Devillaine, C. Navarrete, and D. Silva, "Validating travel behavior estimated from smartcard data," *Transportation Research Part C: Emerging Technologies*, vol. 44, pp. 70–79, 2014.
- [17] T. Kusakabe and Y. Asakura, "Behavioural data mining of transit smart card data: a data fusion approach," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 179–191, 2014.
- [18] X.-L. Ma, Y.-H. Wang, F. Chen, and J.-F. Liu, "Transit smart card data mining for passenger origin information extraction," *Journal of Zhejiang University SCIENCE C*, vol. 13, no. 10, pp. 750–760, 2012.
- [19] J. Zhao, A. Rahbee, and N. H. M. Wilson, "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 376–387, 2007.
- [20] E. Fishman, "Bikeshare: a review of recent literature," *Transport Reviews*, vol. 36, no. 1, pp. 92–113, 2016.
- [21] M. Ahillen, D. Mateo-Babiano, and J. Corcoran, "Dynamics of bike sharing in Washington, DC and Brisbane, Australia: implications for policy and planning," *International Journal of Sustainable Transportation*, vol. 10, no. 5, pp. 441–454, 2016.
- [22] B. Caulfield, M. O'Mahony, W. Brazil, and P. Weldon, "Examining usage patterns of a bike-sharing scheme in a medium sized city," *Transportation Research Part A: Policy and Practice*, vol. 100, pp. 152–161, 2017.
- [23] R. Beecham and J. Wood, "Exploring gendered cycling behaviours within a large-scale behavioural data-set," *Transportation Planning and Technology*, vol. 37, no. 1, pp. 83–97, 2014.
- [24] M. Bordagaray, L. dell'Olio, A. Fonzone, and Á. Ibeas, "Capturing the conditions that introduce systematic variation in bike-sharing travel behavior using data mining techniques," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 231–248, 2016.
- [25] X. Ma and Y. Wang, "Development of a data-driven platform for transit performance measures using smart card and GPS data," *Journal of Transportation Engineering*, vol. 140, no. 12, Article ID 04014063, 2014.
- [26] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207–216, Washington, DC, USA, 1993.
- [27] B. Keuleers, G. Wets, T. Arentze, and H. Timmermans, "Association rules in identification of spatial-temporal patterns in multiday activity diary data," *Transportation Research Record*, no. 1752, pp. 32–37, 2001.
- [28] B. Keuleers, G. Wets, H. Timmermans, T. Arentze, and K. Vanhoof, "Stationary and time-varying patterns in activity diary panel data: explorative analysis with association rules," *Transportation Research Record*, vol. 1807, pp. 9–15, 2002.
- [29] D. Kusumastuti, E. Hannes, D. Janssens, G. Wets, and B. G. C. Dellaert, "Scrutinizing individuals' leisure-shopping travel decisions to appraise activity-based models of travel demand," *Transportation*, vol. 37, no. 4, pp. 647–661, 2010.
- [30] M. Diana, "Studying patterns of use of transport modes through data mining," *Transportation Research Record*, vol. 2308, pp. 1–9, 2012.
- [31] K. K. A. Chu and R. Chapleau, "Augmenting transit trip characterization and travel behavior comprehension: multi-day location-stamped smart card transactions," *Transportation Research Record*, vol. 2183, pp. 29–40, 2010.
- [32] Y. Zhang, T. Thomas, M. Brussel, and M. van Maarseveen, "Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China," *Journal of Transport Geography*, vol. 58, pp. 59–70, 2017.
- [33] T.G.o. Nanjing, *Planning Standards for Public Facilities*, T.G.o. Nanjing, Nanjing, China, 2015.