Xiaobo Qu
Lu Zhen
Robert J. Howlett
Lakhmi C. Jain *Editors*

Smart Transportation Systems 2020

Proceedings of 3rd KES-STS
International Symposium

Smart Innovation, Systems and Technologies

Volume 185

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, Google Scholar and Springerlink **

More information about this series at http://www.springer.com/series/8767

Xiaobo Qu · Lu Zhen · Robert J. Howlett ·
Lakhmi C. Jain
Editors

Smart Transportation Systems 2020

Proceedings of 3rd KES-STS International
Symposium

Springer

Editors
Xiaobo Qu
Dept of Architecture and Civil Engineering
Chalmers University of Technology
Göthenburg, Västra Götalands Län, Sweden

Lu Zhen
Management Science and Engineering
Shanghai University
Shanghai, China

Robert J. Howlett
Bournemouth University
Poole, UK

Lakhmi C. Jain
University of Technology Sydney
Sydney, NSW, Australia

Preface

The transport and mobility technologies are emerging recently with respect to different transport modes ranging from private car transport, public transport, maritime transport, freight transport, and airport transport. The recent advent and progress of communication, and computational and control technologies make it possible to fundamentally shift the transport management schemes in a wide concept. In this regard, the Third International Symposium on Smart Transport Systems is going to be organized in June 2020, in a very difficult time amidst the coronavirus pandemic, to provide a communication and collaborative platform among researchers in the broad area of smart transport. We are contributing our parts to make our society more efficient and robust. Together we can beat this.

For the year of 2020, 18 papers were accepted in the broad area of smart transport systems. Two papers are in the area of travel behaviour modelling, discussing the modal shift and crowding effect. Two papers are focused on maritime transport—port operations and marine traffic emission assessment. Three papers deal with the topic of the environmental impact of transport systems. Four papers are in the area of public transport modelling, ranging from performance evaluation and electric public transport design to incident management and bus bridging. Two papers address the needs triggered by next-generation freight transport and logistics. The rest five papers are focused on traffic flow modelling, especially in the era of connected and automated vehicles that might fundamentally change our transport systems in the future. These papers were rigorously peer-reviewed by at least two independent assessors and one editorial member. We establish a dialogue between assessors and authors in the progress of improving these papers. Participants and authors are mainly scholars and practitioners from Europe, China, Australia, Canada, and the USA.

Gothenburg, Sweden	Xiaobo Qu
Shanghai, China	Lu Zhen
Poole, UK	Robert J. Howlett
Sydney, Australia	Lakhmi C. Jain

Contents

About the Editors

Xiaobo Qu is a Professor and research group leader at the Division of Geology and Geotechnics, Urban Mobility Systems research group, Chalmers University of Technology. Throughout his academic career, he has endeavoured to practically improve transport safety, efficiency, equity, and sustainability through traffic flow modeling, network optimization, and most recently, emerging technologies. In particular, his research has been applied to the improvement of emergency services, operations of electric vehicles and connected automated vehicles, and management of vulnerable road users. He has authored or co-authored over 90 articles published in leading transport engineering journals, and is a recipient of many prestigious awards. His research has been supported by the Australian Research Council Discovery Programme, Queensland Department of Transport and Main Roads, Sydney Trains, National Natural Science Foundation of China, Swedish Innovation Agency Vinnova, and the European Union. Xiaobo is currently an Associate Editor/Editorial Board Member for IEEE Trans. on Cybernetics, IEEE ITS Magazine, Journal of Transportation Engineering ASCE, Transportation Research Part A, Computer-Aided Civil and Infrastructure Engineering, etc.

Lu Zhen is a Professor and Dean in the School of Management at Shanghai University. He got his B.S. and Ph.D. degrees from Shanghai Jiao Tong University in 2003 and 2008, respectively. He worked at National University of Singapore as a postdoctoral scholar before joining Shanghai University in 2011. His research interests include logistics and supply chain management, operations research, optimization in port and shipping management, and knowledge management. He has published 75 SCI and SSCI journals including Transportation Science, Transportation Research Part B, Naval Research Logistics, and IISE Transactions. He is the PI of 15 projects including four at the national level. He has served as an associate editor/editorial board member of four SCI/SSC journals such as Transportation Research Part B (ABS 4), Journal of the Operational Research

(ABS 3), Computers & Operations Research (ABS 3); and he is also the Fellow of the Operational Research Society (U.K.). He has been awarded Changjiang Young Scholar in China.

Dr. Robert J. Howlett is the Executive Chair of KES International, a non-profit organization that facilitates the dissemination of research results in areas including intelligent systems, sustainability, and knowledge transfer. A Visiting Professor at Bournemouth University, UK, his technical expertise concerns the use of intelligent systems to solve industrial problems. He has been successful in applying artificial intelligence, machine learning, and related technologies to sustainability and renewable energy systems; condition monitoring, diagnostic tools and systems; and automotive electronics and engine management systems. His current research focuses on the use of smart microgrids to achieve reduced energy costs and lower carbon emissions in areas such as housing and protected horticulture.

Dr. Lakhmi C. Jain, Ph.D., M.E. Fellow (Engineers Australia), is currently serving at the University of Technology Sydney, Australia, and Liverpool Hope University, UK. Professor Jain also serves at KES International, which provides the professional community with opportunities for publication, knowledge exchange, cooperation, and team-building. Involving over 5,000 researchers drawn from universities and companies worldwide, KES facilitates international cooperation and generates synergies in teaching and research. KES regularly provides net-working opportunities for the professional community through one of the largest conferences of its kind.

A Decision Support System Based on Transport Modeling for Events Management in Public Transport Networks

Luca Studer, Paolo Gandini, Giovanna Marchionni, Marco Ponti, Sergio Arduca, and Serio Agriesti

Abstract This paper presents a modeling approach developed within the MOTUS project, designed to provide a standardized and solid intervention proposal to face events and disruption on a public transport network. This modeling approach resulted into a tool capable of identify in a formalized way the nodes and links where to broadcast info-mobility information through ITS systems and to lead the users to the best alternative solutions. The tool is exploited to make the decision process less dependent on the expert judgment (that still plays a vital role) and human factors, to allow the service provider to respond in a faster and clear way to the possible disruptions both through info-mobility and the strengthening of the offer on the involved routes. Therefore, this paper describes how the modeling approach is applied, how the resulting tool can be exploited, and finally provides an example on the city of Milan, simulating the closure of one of the main lines and reporting the results provided by the presented model and the developed tool.

1   Introduction

The continuous development of collective transport networks in metropolis, along with the application of demand management policies aimed at discouraging the individual motorized mobility, led in the last years to a continuous increase in the use of collective means, both for systematic and for occasional travels. This trend has significant direct and indirect benefits at collective level. For this virtuous process to continue in the near future, the Local Public Transport (hereinafter LPT) system shall

L. Studer (✉) · P. Gandini · G. Marchionni · M. Ponti · S. Agriesti
Mobility and Transport Laboratory, Politecnico Di Milano, Dipartimento Di Design, Milan, Italy
e-mail: luca.studer@polimi.it

S. Arduca
Direzione Tecnica—Unità Coordinamento Manutenzione Impianti, Milan Municipality, Milan, Italy

1

provide high-efficiency standards to satisfy users and their needs as far as possible, e.g., by providing them with a regular and punctual service.

Over the years, the companies managing LPT systems therefore developed operating procedures and solutions to increase the regularity and effectiveness of the service under ordinary service conditions, for example, satellite location systems for vehicles and real-time user information systems. Even though these solutions, as well as other technological innovations, allowed to increase the competitiveness of LPT with respect to private transport under standard service conditions, several criticalities still concern the management of emergencies caused by relevant disruptive events. Inconveniences prove particularly significant in case of service interruption on a main urban line, e.g., a railway or subway section, under consideration of the high number of users they affect.

In this context, a macro-model agile enough to evaluate different interventions and a tool to automatically produce the needed info-mobility strategies were developed to manage the events causing the interruption or the reduction of an LPT service. In fact, the designed model aims at identifying and proposing both interventions on the transport supply (dimensioning of the replacement service, increase in the number of means on the lines near the event, etc.), and strategic indications to manage the demand, which means providing users with focused, timely, and effective information so they can choose more appropriate routes and means. The main contribution of the MOTUS project, as reported in this paper, was exactly to formalize the operating procedures and how they are defined during these disrupting events, so to eliminate or at least reduce the weight of human factors on the managing process. The tool developed within MOTUS is, in fact, flexible enough to be transferred and applied to other urban transport networks, as it will be shown in the following.

The structure of the model and of the corresponding tool for the analysis and management of collective transport under disrupted service conditions follows these steps:

- construction of a simulation model describing the collective transport system and simulating its operation under both standard and disrupted conditions;
- implementation of an analysis and intervention procedure to manage emergencies by enhancing the offer system in case of disruption. Threshold for different relevant indicators was identified so that the service provider would retain control over the adopted strategies while still enjoying the benefits of a semi-automated simulation process;
- implementation of a procedure and definition of the tool to identify strategic network nodes where users should be provided with useful information, in case of disruption, for them to spread out to new itineraries based on the optimum configuration given by the model.

2 State of the Art

At operational level, the management of events causing interruptions, slowdowns, or more general consequences on a LPT network typically relies on the competence and experience of operators and functionaries of an operating center. The model and its tool, as described in this paper, refer to problems that are widely analyzed in the scientific literature and concern disruption management and user information management.

Within disruption management, a recent paper [1] provides an interesting model to identify and characterize disruptions by analyzing users' behavior which, for the purposes of the model, is divided into three categories: missed passengers, detoured passengers, and delayed passengers. The model, which is applied to the metropolitan network of Beijing, allows to quantify the effects of a disruption in terms of journey time and delay.

Disruption management was also historically analyzed in detail in railways, a field offering a wide literature in particular on the timetable rescheduling problem. For example, the most recent papers on the subject focus on the development of methods and models in cases when the management of the transport supply is committed to different objectives: passengers' satisfaction, control of operational costs and of variations to the original time schedule [2], and minimization of delays and cancellations [3]. Again in railways, the paper by [4], identifies and effectively highlights the three variables, railway schedule, staff, and rolling stock management, to intervene in organizational terms to minimize inconveniences in case of disruption, thus developing a model that can enhance the configuration of the variables with few iterations.

A frequently used approach in disruption management, in the emergency planning, refers to the analysis of the vulnerability of transport networks by means of techniques, theories, and models deriving from the theory of graphs. A wide overview and dissertation on this subject is given by Mattsson and Jenelius [5] in their work collecting the most recent researches on the subject, while a current application is presented, for example, in paper [6], which analyzes the robustness of the subway network of Beijing based on the complex network theory.

Other interesting papers on disruption management are [7], which presents a model developed in CPLEX to manage a disruption affecting the LPT of Vienna, and [8], which develops a model tested on the subway network of Naples to define the optimum intervention strategy in case of disruption on railway and metropolitan networks, also considering stochastic variations of train performances and delays.

Useful references for the model that is implemented and described in this paper are also researches and publications referring to methods, models, and theories on the management of private transport networks, in particular [9–13].

Other useful references for the management of passenger rerouting are [14], which addresses the problem of congested train stations along a subway line having boarding/alighting difficulty due to large volumes of boarding/alighting commuters, and [15], which proposes a novel probabilistic methodology to estimate the bus dwell time

taking into account the interactions among buses, arrival passengers, and surrounding traffic.

The effective and ready management of user information in case of disruption is an efficient tool to limit inconveniences. On this subject, a useful reference is the paper by Cats and Jenelius [16] that analyzes the mitigating effect that real-time information may have on the impacts of disruptions in LPT. In order to define strategic links and nodes of a network where users are to be informed, criteria such as the number of users, their needs, and the actual possibilities of the network manager are taken into consideration. Other useful published papers that analyze the topic of user information in detail are [17], which analyzes the benefits obtainable by informing travelers of integrated public transport options using real-time accurate information, and [18], which investigates customers' desired quality of Integrated Multimodal Travel Information (IMTI) provision in public transport. Both papers are based on surveys carried on public transport users in the city of Melbourne and in Holland, respectively. They allow to understand the great importance given by users on the information they receive during their journey, above all in case of disruptions which force them to change their habits.

[19] proposes the design and development of a real-time mobility information system concerning public transportation in the city of Milan, conceived in synergy with the methodology and the tool presented in this paper; however, [19] is mainly aimed at developing a travel planner that can suggest personalized travel solutions for individual users involved in a disruption. Finally, [20] and [21] describe specific modeling aspects that flowed into the general methodology presented in this paper.

3 The Macro-simulation Model and the Disruption Management Tools

The management of emergency situations on an integrated network of collective public transport is a task requiring, for every metropolitan context and in peace time, the development of defined procedures, which must be as far as possible standardized and shared between the several subjects that are to operate in case of relevant events. However, in some contexts or situations, the decisions may only be taken when an event occurs, based on criteria defined by the staff on duty and on their experience. The consequent interventions, although focused on solving the disruption, do not consider the global condition of the network and can hardly comply with criteria of resource optimization and minimization of inconveniences.

A greater effectiveness of emergency management can be obtained by engineering the procedures to follow in case of disruption. The aim of the presented modeling approach is to offer intervention solutions complying both with a precise criterion of optimization of the available resources and with the minimization of overall time-wasters at collective level, i.e., all network users.

The implemented model consists of different phases to carry out sequentially. The first stage is a double simulation that allows a comparison of users' flow on the various sections of the network between the Baseline Scenario (mobility as usual) and the Disruption Scenario. By evaluating the results of the first comparison, the model utilizer (generally a control room operator) may define the main parameters governing the following steps of the model. In the successive step, the model determines the necessary enhancements of the offer system during the disruption. In the last phase, another comparative simulation is carried out between the Baseline Scenario and the Intervention Scenario (defined as the Disruption Scenario with the addition of the implemented enhancements); on the basis of this comparison, the model determines the content and the location of the information to be provided to users.

The main contribution developed during the MOTUS project is represented by the intermediate and the last phase: In fact, the designed model is not limited to the simulation of the transport network considering the disrupted segments, but actually grants the service provider with the routes to be strengthen and the needed transport means on the basis of some service thresholds, and determines the necessary information to communicate to users. Figure 1 provides the overview of the process through a flowchart, detailing when in the process the developed tools exploit the simulation outputs.

Simulations aside, this paper also describes the info-mobility and intervention tools, both designed through the Cube software environment and its programming

Fig. 1 Model flowchart

Fig. 2 Extract of the tool script

language. Cube is a modeling platform that covers all aspects related to transportation planning, engineering, and land use. A brief example of the coding of the model is shown in Fig. 2. Through these tools, the control room operator is able to automatically obtain both the lines to be strengthened, the number of additional vehicles and the messages to be broadcasted at each node.

3.1 First Simulation: Normal State of the Network (Baseline Scenario)

This first phase does not differ much from macro-simulations as found in literature, but it is necessary to frame the state of the network and define a benchmark against which compare the effects of the disruption and the goodness of the adopted interventions (on which the info-mobility tool defines the messages to be broadcasted and where to broadcast them). Still, it is worth to highlight that the model was calibrated with a focus on the LPT routes on the basis of the travel times; the ones obtained through the model during the application to the Milan case study were compared with the ones provided by ATM (the transport company managing the public transport lines in Milan).

Moreover, the model adopts the hyperpath algorithm; the process is based on proper parameters and consists in two main phases: the identification of hyperpaths and the calculation of the probability to choose each route. In the first phase of hyperpath identification, the model identifies a discrete set of possible routes for each origin–destination pair. Theoretically, this process may generate an infinite number of hyperpaths. In order to avoid this, the following approach is applied:

- For each origin–destination pair, the minimum-cost hyperpath is identified.
- Further possible hyperpaths are found that have a lower cost than a limit value that is specifically defined.
- At last, identified routes are eliminated if they show a number of interchanges that are higher than a specified value.

In the second phase of the process of calculation of routes, the probability to choose each route is defined.

3.2 Second Simulation: Disruption Scenario and Scenarios Comparison

The above-mentioned process of assignment, carried out for the normal (non-disrupted) state of the network, is replicated for the disrupted condition of the network. The events of disruption are given to the model as inputs files (through .csv files for surface lines and through a dashboard specifically developed for high-capacity lines like subways) containing all the disrupted lines, the involved direction, the involved nodes, and the kind of event (closure of line segment or closure of a station).

The two simulations (Baseline Scenario and Disruption Scenario) are compared to highlight the different distribution of passenger flows along the network. As a first output at the end of this simulation phase, the model generates, for each LPT line, a list containing the flow/capacity value for each link of the line in both scenarios (in Fig. 3 is reported an extract referred to the M3 subway line of Milan).

In addition, the model determines the average variation along the line and the maximum change (which occurs on the section that is mostly affected by the disruption). For example, in Fig. 4 are shown flow/capacity changes (in absolute and percentage) for the lines of the Milan LPT network that has positive variations of these values.

A	B	DIST	VOL_1	FMCAPACITY	RAPP_V_C_1	VOL_2	FMCAPACITY	RAPP_V_C_2
30085	30082	1.68	2970.91	38640	0.08	2970.39	38640	0.08
30082	30077	0.79	8872.67	38640	0.23	8906.93	38640	0.23
30077	30072	0.67	10194.74	38640	0.26	10226.72	38640	0.26
30072	30068	0.37	11275.17	38640	0.29	11305.45	38640	0.29
30068	30063	0.86	12580.43	38640	0.33	12610.67	38640	0.33
30063	30057	0.82	13311.24	38640	0.34	13491.28	38640	0.35
30057	30050	0.66	14100.54	38640	0.36	14334.69	38640	0.37
30050	30041	0.78	14123.91	38640	0.37	14417.02	38640	0.37
30041	30042	0.43	14677.51	38640	0.38	15852.8	38640	0.41
30042	30046	0.69	9018.12	38640	0.23	12897.44	38640	0.33
30046	30048	0.55	8300.84	38640	0.21	12184.87	38640	0.32

Fig. 3 Flow/capacity ratio in Baseline Scenario (1) and Disrupted Scenario (2) for each link of M3 subway line of Milan. Link with maximum flow/capacity ratio is highlighted

NOME_LINEA	RAPP_MED_1	RAPP_MED_2	D_RAPP_MED	P_RAPP_MED	RAPP_MAX_1	RAPP_MAX_2	D_RAPP_MAX	P_RAPP_MAX
TRAM33	0.14	0.24	0.1	71.43	0.31	0.65	0.34	109.68
TRAM7	0.06	0.08	0.02	33.33	0.16	0.23	0.07	43.75
Bus48	0.1	0.13	0.03	30	0.2	0.26	0.06	30
Bus37	0.19	0.25	0.06	31.58	0.58	0.74	0.16	27.59
M1	0.26	0.3	0.04	15.38	0.57	0.72	0.15	26.32
TRAM2	0.34	0.4	0.06	17.65	0.86	1.03	0.17	19.77
TRAM14	0.37	0.38	0.01	2.7	0.72	0.86	0.14	19.44
TRAM3	0.25	0.3	0.05	20	0.67	0.8	0.13	19.4
TRAM15	0.21	0.24	0.03	14.29	0.56	0.66	0.1	17.86
Bus965	0.49	0.55	0.06	12.24	1.4	1.64	0.24	17.14
M3	0.16	0.18	0.02	12.5	0.38	0.44	0.06	15.79
TRAM23	0.25	0.28	0.03	12	0.55	0.63	0.08	14.55
Bus79	0.06	0.05	-0.01	-16.67	0.14	0.16	0.02	14.29
Bus713	0.08	0.09	0.01	12.5	0.22	0.25	0.03	13.64
BUS190	0.1	0.1	0	0	0.22	0.25	0.03	13.64
S11	0.89	0.94	0.05	5.62	2.14	2.43	0.29	13.55
Bus70	0.08	0.08	0	0	0.26	0.29	0.03	11.54

Fig. 4 Average (med) and maximum (max) flow/capacity variation in absolute value (d) and percentage (p) for some Milan LPT lines. Line with maximum percentage increase is highlighted

The analysis of the results given, in terms of extent of the variations on the several lines, allows to quantify the effects of the disruption on the public transport system and the priorities of any intervention.

3.3 Intervention Tool: Strengthening of the Service and Intervention Scenario

In order to define the strengthening of the service to be implemented, some criticality thresholds related to flow/capacity values are defined, beyond which a line can be defined as critical. In particular, a generic line l is considered critical if the following three conditions are meet:

- The absolute value $R^D_{l\,\max}$ for the lth line of the maximum flow/capacity ratio in the Disruption Scenario exceeds a predetermined threshold $\overline{R}_{\min}$:

$$R^D_{l\,\max} > \overline{R}_{\min} \tag{1}$$

- The percentage increase $\Delta P_{l\,\max}$ for the lth line of the maximum flow/capacity ratio between the Baseline Scenario and the Disruption Scenario is greater than a predetermined threshold $\overline{P}_{\min}$:

$$\Delta P_{l\,\max} > \overline{P}_{\min} \tag{2}$$

- The ratio between the linear extension of the lth line sections characterized by critical conditions (as defined by the previous two points) $l_{\mathrm{CRIT}\,l}$ and the total line length L_l is greater than a predetermined threshold $\overline{E}_{\min}$:

$$\frac{l_{\text{CRIT}\,l}}{L_l} > \overline{E}_{\min} \tag{3}$$

The first two conditions refer to single line links. The simultaneous exceeding of threshold (1) and threshold (2) makes the line link critical. If the number of critical links is such as to affect a percentage extension of the total line route that is higher than the threshold (3), the line itself will be considered as critical and therefore in need of enhancement.

The setting of the thresholds values depends on the overall capacity of the LPT operator to cope with offer enhancements and can therefore be calibrated by simulating a set of possible disruptions on the network.

In particular, condition (1) allows to define an absolute threshold lower limit; ideally, the value of $\overline{R}_{\min}$ should be set at 1 or even less than 1 if to ensure a greater level of comfort. Contrariwise, in case of a network normally very congested, during emergency conditions the flow/capacity ratio can exceed the value of 1 on several links and several lines. In this case, considering the limits of the resources available (means and people), and requiring intervention priorities to be defined, it is possible to set the $\overline{R}_{\min}$ threshold even above the unit value, by calibrating it on a value that is considered appropriate by the LPT network operator.

The fulfillment of condition (2) guarantees that only those links, and potentially the lines, where crowding considerably increases because of the disruptive event will be considered as critical. Lines that are already overcrowded in standard conditions are therefore excluded. A decrease in crowding on these sections should hopefully be obtained in the long term, but it is not relevant for the aims of the emergency management which this model is implemented for.

Condition (3) sets a limit on the extension of the criticality along a LPT line to make it require an enhancement: Service enhancements are only planned on a line that is considered as critical based on conditions (1) and (2), if it involves a higher percentage of the total route than the set threshold.

The developed tool allows the dynamic definition of threshold values requiring service enhancements or not. As a function of the single case study, of the significance of the disruptive event and of the availability of mitigating resources, it is therefore possible to vary these thresholds to effectively react to the emergency. If the real-time model is used to manage emergencies, the need to quickly set appropriate values requires an adequate sensitivity in use on the part of operators. Intervention scenarios can nevertheless be prepared in peace time, possibly making iterations with several values and opting for those which most satisfy the operational needs of the transport company. Still in the following case study, the values adopted for the MOTUS project and the city of Milan will be reported.

For the lines defined as critical according to the procedure described above, the model will suggest enhancements. The model is designed in a way that, through the comparison of the Baseline and of the Disruption scenarios, by exploiting the defined thresholds the interventions are automatically obtained. In fact, this phase of the model generates the list of interventions to implement in order to mitigate the effects of the disruptive event. For critical lines, the number of additional means

required is defined; the enhancement will be such as to bring the crowding condition on the line below the threshold value $\overline{R}_{min}$ that initially determined the condition of criticality.

Once the lines to enhance according to the above-mentioned algorithm are defined, the graph of the network is updated with the planned enhancements. The successive simulation consists in the assignment of the demand to the graph modified from the supply side, and the result is the final users' optimum distribution on the network, named Intervention Scenario.

The management of the emergency shall aim at letting users spread out according to the optimum configuration defined by the assignment of the Intervention Scenario, thanks to the information they received along the route as defined by the model itself.

The information, as well as the strategic nodes where it is to be provided, are defined by means of algorithms based on the comparison between the flows assigned in the Intervention Scenario and the flows of the non-disrupted Base Scenario.

Figure 5 shows the result of the process of comparison. LPT lines sections where, in case of disrupted service, users should be redirected and consequently flows will increase are highlighted in green. Conversely, lines where users will decrease are highlighted in red. Lines to reroute users can therefore also be found at qualitative level.
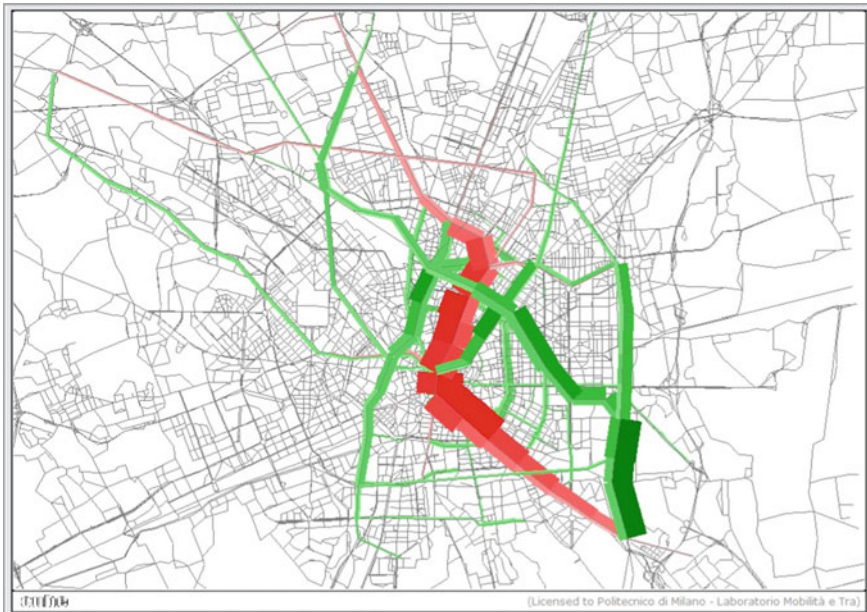
Fig. 5 Comparison between volumes assigned in the baseline scenario and in the intervention scenario in the city of Milan, downstream offer enhancements

3.4 Info-Mobility Tool: Definition of Information to be Provided to Users

Based on the analysis of the way the flows distribution varies with respect to the undisrupted Baseline Scenario, the info-mobility tool will generate a word-processed report containing a list of the aggregate information to provide users with, as well as of the strategic network nodes where it shall be given. This kind of information aims at addressing users' flows toward the best alternative route by means of information that is as simple and immediate as possible.

The implemented algorithm is based on the evaluation of variations (between Baseline and Intervention scenarios), both absolute and in percentage, of the users' flow on the different links of the network. The set of links on which it is necessary to provide information is determined on the basis of two simultaneous conditions:

- The absolute increase or decrease ΔF_{il} in the flow of users on each ith link of the lth line between the Baseline Scenario and the Intervention Scenario is greater than a predetermined threshold $\overline{F}_{\min}$:

$$|\Delta F_{il}| > \overline{F}_{\min} \tag{4}$$

- The percentage increase or decrease $\Delta F_{\%il}$ in the flow of users on each ith link of the lth line between the Baseline Scenario and the Intervention Scenario is greater than a predetermined threshold $\overline{F}_{\%\min}$:

$$|\Delta F_{\%il}| > \overline{F}_{\%\min} \tag{5}$$

Focusing on the set of links identified by the above described conditions, a subsequent condition is placed on the starting and final nodes of the links themselves, in order to definitively identify which are the essential points of the network where to provide information to users (directing them toward their ideal alternative path): The absolute increase or decrease in the numbers of users getting on (ΔON_{ni}) and getting off (ΔOFF_{ni}) at the nth node of the ith link between the Baseline Scenario and the Intervention Scenario is greater than the previously determined threshold $\overline{F}_{\min}$:

$$|\Delta ON_{ni}| > \overline{F}_{\min} \tag{6}$$

$$|\Delta OFF_{ni}| > \overline{F}_{\min} \tag{7}$$

As with the evaluation of enhancements, the thresholds determining the significance of a variation in terms of information diffusion are also defined as input by the evaluator using a specific user interface developed with the tool.

The developed model can specify the text of the information as well as its space location, i.e., its stop node where the message is to be transmitted. Moreover, the

needed information are prioritized on the basis of the number of reached users for each node. The duration of the simulation process is limited so that information is given quickly to prepare staff and means for an enhancement of the service.

4   The Case Study

The developed model and tool were then applied to the Milan city to evaluate different scenarios in which one or more sections of the underground lines can be interrupted; as an example, a specific case is presented below, in which a section of the M3 subway line is interrupted between the stations of Sondrio and Porta Romana, as shown in Fig. 6.

   The simulations were carried out considering the morning peak hour; seven stations were made unavailable due to said closure. Again, it should be highlighted that the results that will be reported in the following were obtained in a semi-automated and formalized way. The value of the developed modeling tool presented in this paper lies, in fact, in the easiness and responsiveness of the model that allows the service provider to obtain both the interventions and the crucial points where to broadcast the info-mobility information. Therefore, it should be clear to the reader that the following results could be obtained for any other disruption of the public transport network, in short times and through a user-friendly tool.
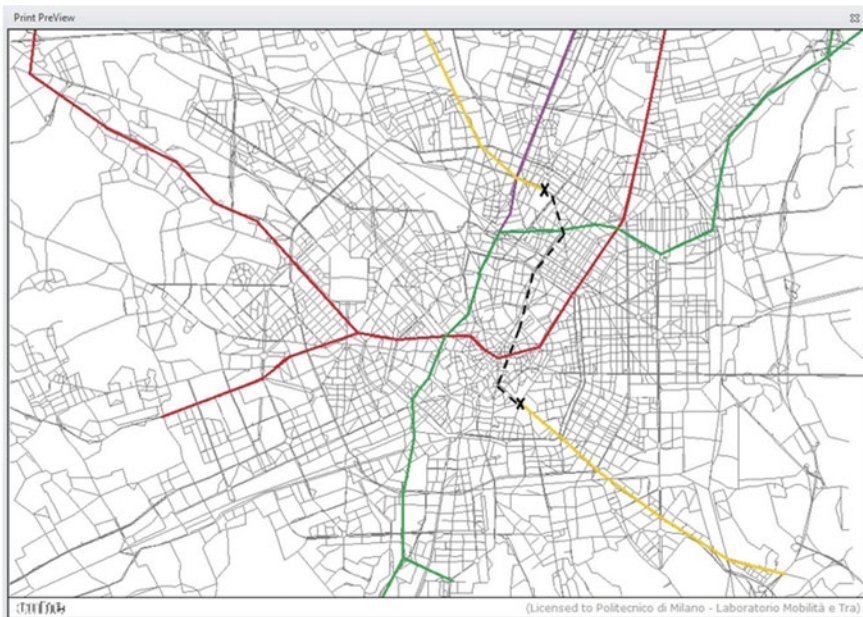
Fig. 6 Section closure on the subway M3 line between Sondrio and Porta Romana

Table 1 Value of threshold used in the Milan case study

Threshold	Brief description	Value
$\overline{R}_{\min}$	Flow/capacity ratio	0.8
$\overline{P}_{\min}$	Percentage increase of flow/capacity ratio	15%
$\overline{E}_{\min}$	Extension of criticalities along the line	20%
$\overline{F}_{\%\min}$	Percentage increase or decrease of passengers' flow on links	20%
$\overline{F}_{\min}$	Absolute increase or decrease of flow on links and of passengers getting on/off at stops	200 (metro, train) 100 (bus, tram)

For the Milan case study, and in particular for the specific scenario presented, the values of the thresholds introduced in Chap. 3 and used in input for the model are detailed in Table 1.

It should be clear that when talking about increase and decrease, it is meant a comparison between the Baseline Scenario and the Disruption/Intervention Scenario.

The intervention tool of the model, based on the first three thresholds of Table 1, suggests for the case study in question the enhancement of one bus line (line 222 with three additional vehicles) and three train lines (one more run for each of them) in their urban section.

The last two thresholds are conversely important for the info-mobility tool: In fact, through these values the tool decides what are the nodes worth of involving in the broadcasting of the information. To use the delta in volume at nodes guarantees that the messages are broadcasted only where and when needed (Fig. 7).

Again, the information tool of the model gives an output that includes both the nodes where to broadcast the information and the set of information itself; an extract of this list of information is shown in Table 2.

Both the nodes of broadcasting and the messages as defined above are given back by the developed model, based on the functional thresholds as defined by the service provider. This allows the public transport service provider to test different events and situations in a flexible and replicable way, without consuming time and efforts. In the same way, the model gives back also the means of transportation and their location to be deployed to restore the service level under the defined thresholds, as will be shown in the following. It should be highlighted that the info-mobility messages are already including the alternative routes and vehicles. Again, should the number of additional vehicles be too high as an effort, the service provider can access the tool and act on the level of service to tune the suggested solutions. Finally, even though in this case study no additional vehicle is suggested for the train line, the line is considered in the tool and can be strengthened as well.
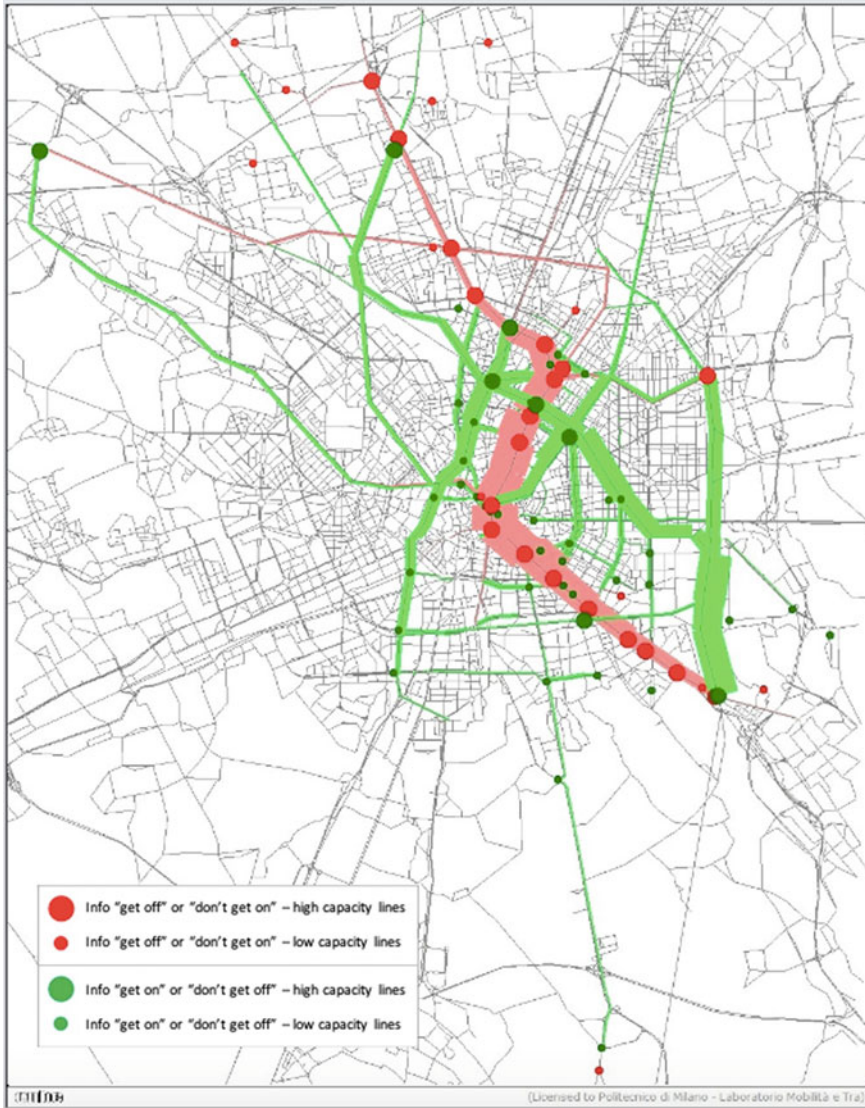
Fig. 7 Output of the info-mobility tool: localization of network points where to provide information to users

5   Conclusions

The aim of the paper was to describe the developed model and highlight how, through the programmed tools, it grants to service providers and public transport companies both the possibilities of planning ahead for certain disrupted scenarios and of reacting

Table 2 Extract of info-mobility messages to be provided to users, as given by the tool

Stop/node (line)	Point of broadcasting	Info-mobility message
Maciachini M3	Surface lines (on-board)	*Don't get off tram line 4 but continue to Lanza M2 - Cairoli M1*
	Within the metro station and at the surface transit stops	*Don't get on line M3*
		Take the tram line *4 up to Lanza M2 - Cairoli M3*
Zara M3	Metro (on-board)	Exit the M3 line
		Don't get off line M5 but continue to Garibaldi FS
	Within the metro station and at the surface transit stops	*Don't get on line M3*
		Take line M5 and continue to Garibaldi FS
Sondrio M3	Metro (on-board)	*Get off line M3*
	Surface lines (on-board)	*Don't get off bus line 43 but continue to Gioia M2 - Turati M3*
	Within the metro station and at the surface transit stops	*Don't get on line M3*
		Take bus line 43 and continue to Gioia M2 - Turati M3
		Take bus line 90 and continue to Centrale M2-M3 - Caiazzo M2 - Loreto M2-M3
		Take bus line 92 and continue to Centrale M2-M3 - Caiazzo M2 - Lodi M3
Centrale M2-M3	Within the metro station and at the surface transit stops	*Don't get on line M3*
		Take tram line 5
		Take tram line 9 and continue to P.ta Romana M3 - P.ta Lodovica - P.ta Genova M2
		Take bus line 92 and continue to Sondrio M3 - Zara M3-M5

promptly and in a formalized way to unforeseen events. Through the description of the case study, it was shown how the tool calculates the best strengthening actions and routes and also identifies the crucial nodes and the information to be broadcasted.

The main highlights are focused on the methodological part; however, it is described through the paper how the model is designed to grant control to the evaluators, through thresholds definition and a flexible simulation activity based on the comparison with the baseline state of the network. The simulation phase of the model does not differ much from a macro-model analysis, so a certain degree of simplicity was maintained in its description; however, the developed tools are based on the outputs of this simulation activity (specifically on the scenario comparisons) and were

designed in a way that presents the results as coherent and coexistent with the results of the simulations themselves.

Therefore, the paper tries to describe and report both the methodological approach and the resulting tools in order to promote as best practice this formalized analysis of public transport networks responding to disruptive events, both planned and unforeseen. It is acknowledged both in industry and in literature that public transport will see growing both its importance and the resulting challenges in future cities and metropolis such as Milan; therefore, the paper tries to contribute with the tools to improve the resilience and the operational performance of the public transport system.

References

1. Sun, H., Wu, J., Wu, L., Yan, X., Gao, Z.: Estimating the influence of common disruptions on urban rail transit networks. Trans. Res. Part A. (2016)
2. Binder, S., Maknoon, Y., Bierlaire, M.: The multi-objective railway timetable rescheduling problem. Trans. Res. Part C (2017)
3. Veelenturf, L.P., Kidd, M.P., Cacchiani, V., Kroon, L.G., Toth, P.: A railway timetable rescheduling approach for handling large-scale disruptions. Trans. Sci. **50**(3), 841–862 (2016). https://doi.org/10.1287/trsc.2015.0618
4. Dollevoet, T., Veelenturf, L.P., Wagenaar, J.C., Huisman, D., Kroon, L.G.: Application of an iterative framework for real-time railway rescheduling. Comput. Oper. Res. (2016)
5. Mattsson, L.G., Jenelius, E.: Vulnerability and resilience of transport systems—a discussion of recent research. Trans. Res. Part A (2015)
6. Yang, Y., Liu, Y., Zhou, M., Li, F., Sun, C.: Robustness assessment of urban rail transit based on complex network theory: a case study of the Beijing subway. Saf. Sci. (2015)
7. Kiefer, A., Kritzinger, S., Doerner, K.F.: Disruption management for the Viennese public transport provider. Public Transport. (2016)
8. Botte, M., D'Acierno, L., Montella, B., Placido, A.: A stochastic approach for assessing intervention strategies in the case of metro system failures. In: AEIT (2015)
9. Russo, F., Vitetta, A.: Metodi per l'analisi dei sistemi di trasporto in condizioni di emergenza. In: Angeli, F. (ed), Milan (2004)
10. Velonà, P., Vitetta, A.: Analisi e gestione della domanda di una rete di trasporto in condizioni di emergenza mediante procedure di assegnazione pseudo-dinamica. In: Angeli, F. (ed), Milan (2004)
11. Russo, F.: Un approccio dinamico alla modellizzazione della scelta del percorso nei sistemi di trasporto collettivo in presenza di variazioni interperiodali dell'offerta. F. Angeli, Milan (2001)
12. Studer, L., Marchionni, G., Arditi, R., Maja, R., Ponti, M.: Trans-Alps: a transport model supporting traffic management in the Alps. In: VI th European Congress and Exhibitions on Intelligent Transport Systems and Services. Aalborg, 18–20 June (2007)
13. Studer, L., Marchionni, G., Ponti, M.: Evalutation of travel and traffic information between Italy and Slovenia in the Promet Project. In: ISEP 2009—Sustainable Transport and Mobility. 26/3/2009, Ljubljana, pp. R61–R64
14. Wang, S., Zhang, W., Qu, X.: Trial-and-error train fare design scheme for addressing boarding/alighting congestion at CBD stations. Transp. Res. Part B **118**, 318–335 (2018)
15. Meng, Q., Qu, X.: Bus dwell time estimation at bus bays: a probabilistic approach. Transp. Res. Part C **36**, 61–71 (2018)

16. Cats, O., Jenelius, E.: Dynamic vulnerability analysis of public transport networks: mitigation effects of real-time information. Netw. Spat. Econ. (2011)
17. Cebon, P., Samson, D.: Using real time information for transport effectiveness in cities. City Cult. Soc. (20110
18. Grotenhuis, J., Wiegmans, B.W., Rietveld, P.: The desired quality of integrated multimodal travel information in public transport: customer needs for time and effort savings. Transp. Policy (2007)
19. Bruglieri, M., Bruschi, F., Colorni, A., Luè, A., Nocerino, R., Rana, V.: A real-time information system for public transport in case of delays and service disruptions. Transp. Res. Procedia **10** (2015)
20. V&B Software Services Ltd.: Sviluppo di un modello del trasporto pubblico della città di Milano, finalizzato all'analisi dell'impatto di "emergenze" sul sistema. CITILABS
21. Arduca, S.: Metodologie e strumenti modellistici a supporto della gestione del trasporto pubblico locale in caso di eventi perturbativi rilevanti: ottimizzazione dei flussi e gestione delle informazioni all'utenza. Politecnico di Milano, Milan (2015)

Existing and Future Investigation of Charging Technology for Electric Bus

Ziling Zeng, Danni Cao, and Xiaobo Qu

Abstract Bus fleet electrification achieves momentum and inspiration within public transport aiming at further improving the mobility sustainability. In many countries, such as Sweden, China, and the USA, there are several ongoing demonstration projects of electric buses and many research projects. The charging technology development and implication is key for the expansion of electric buses and to foster it. An investigation of characteristics and benefits of various existing and future charging technologies has been created in this paper. The main types of charging infrastructure are depot charging, station charging, and inductive charging. The choice of different types is highly related to infrastructure construction, investment, and daily operation. The detailed illustration and analysis of them can provide a solid foundation to the near-future large-scale electric buses' operation.

Keywords Electric bus · Charging technology · Depot charging · Station charging · Inductive charging

1 Introduction

Recently, researches proposed an estimation of the benefit of bus fleet electrification showing that a larger number of electric vehicles can synergistically deliver greater air quality, climate, and health benefits. For example, when electric bus accounts for 27% of the total bus fleet, it can largely reduce the annual concentrations of NOx, O_3, and PM2.5 (in China). Furthermore, the number of annual premature deaths

Z. Zeng (✉) · D. Cao · X. Qu
Department of Architecture and Civil Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden
e-mail: ziling@chalmers.se

X. Qu
e-mail: xiaobo@chalmers.se

D. Cao
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, 100044 Beijing, China

caused by air pollution can be reduced by 17,456 [1]. Meanwhile, some projects with electrified buses are carried out. In Gothenburg, Sweden, 300 electric buses are newly bought and will be implemented in the near future. In Europe, Zero Emission Urban Bus System (ZeEUS) project studies different electrical systems for urban buses including depot charging, station charging, and inductive charging. The construction and maintenance fee are important when deciding which system to apply. These costs vary from different type of batteries, chargers, and the detailed system design. In this case, the main challenge is to figure out the advantages and disadvantages of different charging strategies, in order to provide a cost-effective system.

The main barrier in the rapid development of electric bus services in Europe is the infrastructure-related problem. First, the cost of constructing on-board large capacity batteries for depot charging is extremely high. Second, for station charging, it lacks dedicated infrastructure and robust charging station deployment plan, and it requires high infrastructure and bus acquisition costs [2]. Balancing the operation cost and charging requirements is crucial to provide a more efficient and cost-effective bus operation system.

There are mainly two concepts for the charging of batteries, standard and fast charging. Standard charging is adapted mainly in the bus depot overnight and during longer brakes with a moderate charging power. This causes a high battery capacity and a high weight of the system, since electric buses serve routes during the entire day. Fast charging on the track during operation can reduce the battery capacity and more importantly reduce the weight significantly. However, the bus schedule should provide enough buffer time for charging at certain locations. Especially inductive solutions offer the possibility to charge during driving, but this has not yet been implemented for a public electric bus due to the lack of an integrated charging and serving plan.

Terminal stop charging and along-route charging can be combined as station charging. Compared to depot charging, it provides more charging opportunities. However, the idea of inductive charging provides the easiest access and the most effective way for maintaining the battery in an optimal range of electricity usage along the route. When deciding which strategy to take, the consideration of energy consumption, charging times and possibilities, operation safety, and battery lifetime should be inevitable.

2 Depot Charging

Depot charging is the most time-consuming charging strategy for electric buses. When buses finish their scheduled routes or stay in the depot during the shift, this strategy is adopted as shown in Fig. 1. This charging is usually overnight or sometimes within the dwell time period with slow chargers (typically 40–120 kW). The full charge process for depot charging takes around 4 h. Fast chargers can also be adopted

Fig. 1 Depot charging process [3]

in overnight charging, sharing the same infrastructures as station charging technology. But currently, depots prefer to be equipped with lower charger considering the construction fee and unlimited charging time.

Centralized and decentralized depot charging scheduling researches raised recently for small-, medium-, or large-scale electric buses considering different constraints, such as battery aging cost, grid distribution, and battery electro-thermal [4–9]. A centralized charging process is managed by a central controller (Fig. 2), while decentralized charging process is operated by individual providers considering personal charging profiles. Figure 2 illustrates a centralized charging system, where AC/DC module converts input AC power into adjustable output DC power and DC/DC module converts a source of direct current (DC) from high voltage level to a low level which is suitable for electric bus. These two modules are monitored by the center controller maintaining the conversion infrastructure and communicating with the electric buses to perform the charging plan according to standardized communication protocol [10].
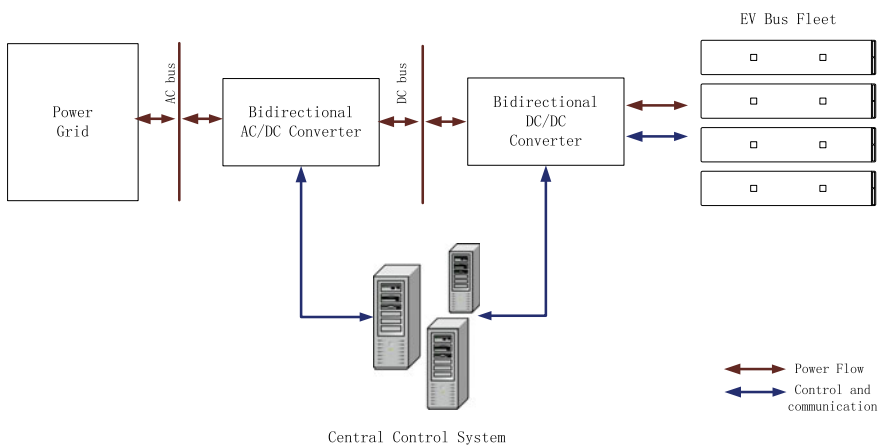
Fig. 2 Centralized depot charging system

From the cost-effective aspect, charging at depot is the one that requires the least amount of infrastructure, as no other equipment is needed except the depot charger. But a large battery capacity is highly concerned, since electric buses need to serve a scheduled route during the daytime without being recharged. However, in some daily operational cases, it is difficult to complete the entire trip without charging. In order to increase the battery capacity, turning bigger and heavier is inevitable, which accordingly increases the overall consumption and reduces the maximum payload of the bus. Although the cost for infrastructure is the least, a large expenditure is required for large-capacity batteries.

From operational optimization aspect, depot charging strategies will not improve current electric bus route plan, since buses are supposed to arrive at depot after serving the whole assigned trips. During the scheduling process, there is no delay caused by charging. If operators aim at improving the scheduled trips, some methodologies can be proposed to manage overnight charging of an electric bus fleet by identifying optimal charging strategy that minimizes the battery aging, charging cost, or maintenance cost [11].

From the grid network aspect, this strategy avoids peak hour charging, where the subscribed power and the maximum charging power delivered by the charger are rather stable. Some optimization algorithms [11] can attribute an optimal charging power for each bus.

3   Station Charging

Station charging refers to bus charging at a certain station within its operational time. Based on the charging location, station charging can be classified as termical stop charging and along-the-route charging.

3.1   Terminal Stop Charging

Terminal stop chargers are placed in the initial or end stops of a bus trip; this placement can be illustrated as shown in Fig. 3, using mainly the regular dwell time at the certain stops to charge their batteries [12]. This charging plan is in a strong linkage between the electric bus scheduling [13] and the charging infrastructure planning. The battery capacity of the bus should be big enough to allow several missed charges, to avoid the bus running out of energy due to external factors such as congestion or emergency. Lower battery capacity requirement reduced the cost of battery, while fast chargers will increase the expenses on station infrastructure.

Terminal stop charging has a rather long dwelling time, so that delays can be compensated, and practical bus running can be adjusted to scheduled time during the charging process [12]. One of the main aspects that need to be considered in this strategy is the impact that charges may have on the schedule. The charging time should
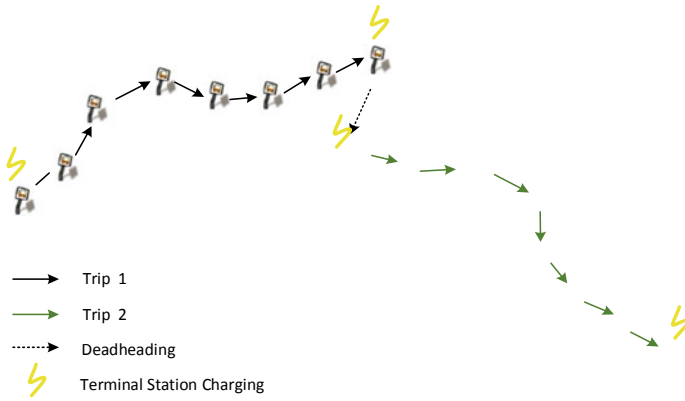
Fig. 3 Terminal stop charging strategy

not exceed the next trip's departure time to avoid trip delay. For scheduling of electric bus terminal stop charging, some solving algorithms such as genetic algorithm [14], dynamic programming [15], exponential smoothing model [16], and locally optimal scheduling [17] techniques have been used in literature.

3.2 Along-the-Route Charging

Along-the-route charging strategy charges electric buses at several intermediate stations along the route as shown in Fig. 4. Two main problems in this system are the deployment of chargers and the charging plan for electric buses.
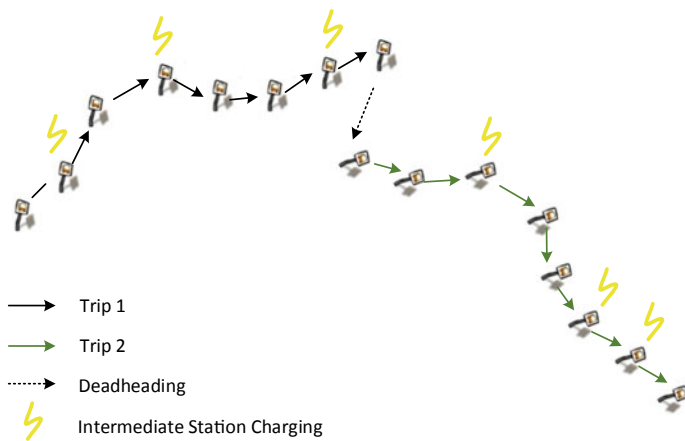
Fig. 4 Along-the-rout charging strategy

For charging infrastructure deployment, the major question is how to plan the deployment and operations of charging facilities along the routes to meet the ever-growing electric bus demand in a systematic and integrated way, and how to couple the traffic and power grid networks. The methodologies will include developing an equilibrium framework, which can be user equilibrium [18] or system equilibrium, capturing the balance between estimated traffic flow and grid power distribution [19, 20]. Based on this framework, a number of chargers along the routes can be allocated, and locations of allocated charging stations and their corresponding capacities can be determined relatively. A trade-off between the capacity of the battery and the amount of charging locations (number of chargers) is also required.

When designing charging plan, the charging cost, energy density, power density, and battery lifetime should be considered [21]. The lifetime of the battery mainly depends on the cycle between charging and discharging. State of charge (SOC) is a measurement of the level of charge of the battery. A fully loaded battery has SOC of 100%, while a fully discharged battery has 0%. To generate an efficient charging plan, the difference between the allowed maximal and the minimal SOC should be kept in a certain optimal level as shown in Fig. 5.

In this strategy, electric bus will be charged according to the current SOC. A frequent charging plan allows smaller batteries, while higher power is needed along the route. The trade-off between battery size, charging power, and different charging strategies can be shown in Fig. 6.

As shown in Fig. 6, the battery capacity for along-the-route charging is the lowest with the lowest battery cost. However, the cost of charging infrastructures and daily operation is the highest among all mentioned strategies.

Fig. 5 Battery state of charge (SOC)
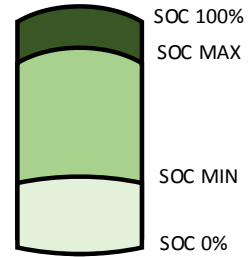
Fig. 6 Requirement for different charging frequency
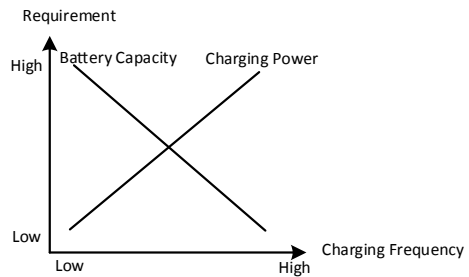
The main benefit of along-the-route charging is the possibility to fully charge electric buses without any interruption during the daily operation. An optimal charging plan will maintain the battery in a certain SOC without letting the batteries deplete completely. Therefore, with this charging strategy, electric bus operation resembles the current diesel buses' operation, and some existing operational strategies can be easily adapted to manage the electric bus fleet [22].

4 Inductive Charging

Using a technology of inductive charging, electricity is transmitted through an air gap from one magnetic coil in the charger to a second magnetic coil fitted to the car as shown in Fig. 7. Accordingly, this wireless system includes two parts—the charger which sits on the floor at the redefined place and a power receiver which attaches to the bus. The charging process can be divided into two classes—charging while parking (static way) or charging while driving (dynamic way). For the dynamic way, the estimation of boarding congestion [23] and bus dwell time [24] at each station should be considered to further avoid delay.

The cost of inductive charging infrastructure is much higher than conductive chargers. A reasonable cost for an inductive charger with capacity to transfer up to 200 kW can be estimated to be 3 MSEK, including on-board pick-up system and power electronics [25]. The corresponding cost for a 300-kW conductive charger, according to the same reference, is estimated to be 1.5 MSEK.

However, inductive wireless charging systems require ferrite cores for magnetic flux guidance and shielding, which are bulky and costly. Also, to control the minimum loss in the ferrites, the charging system is kept under 100 kHz. In this situation, larger coils are needed, and lower power transfer densities occur. The high cost and low power transfer density are particularly problematic for implementing dynamic

Fig. 7 Illustration of inductive charging for buses
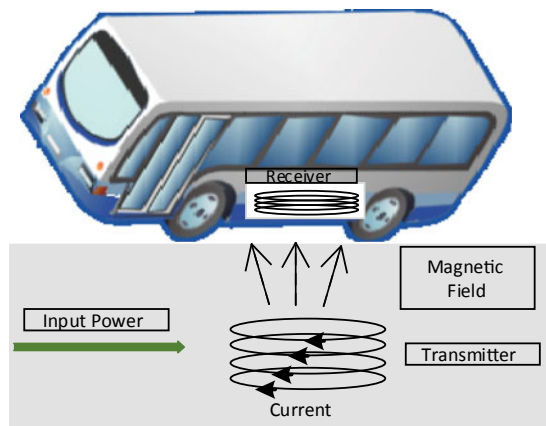
wireless charging, especially for dynamic charging, as the charger should be equipped with a high power capability to deliver enough energy to the electric bus during its very brief time passing over a charging coil [26, 27]. Therefore, this charging strategy has not yet been implemented.

5   Conclusion

Depot chargers are equipped with slow chargers (typically 40 to 120 kW), while station charging is for bus stops (up to 600 KW) or terminals (usually between 150KW to 500 KW) using conductive or inductive chargers. Station charging requires less energy to be stored in the bus, which could significantly reduce the capital costs. However, the construction expenses are much higher compared to depot charging. Wireless charging requires the costliest infrastructure and has the lowest power transfer density, while it enjoys the least requirement of battery capacity and the highest guarantee of battery health. Recently, wireless charging is yet to become commercially viable, although a few experimental systems have been demonstrated.

References

1. Liang, X., Zhang, S., Wu, Y., Xing, J., He, X., Zhang, K., Hao, J.: Air quality and health benefits from fleet electrification in China. Nature Sustain. **2**, 962–971 (2019)
2. Xylia, M., Leduc, S., Patrizio, P., Kraxner, F., Silveira, S.: Locating charging infrastructure for electric buses in Stockholm. Transp. Res. Part C Emerg. Technol. **78**, 183–200 (2017)
3. Ebusco Homepage. https://www.ebusco.com/charging
4. Marongiu, A., Roscher, M., Sauer, D.U.: Influence of the vehicle-to-grid strategy on the aging behavior of lithium battery electric vehicles. Appl. Energy **137**, 899–912 (2015)
5. Schoch, J., Gaerttner, J., Schuller, A., Setzer, T.: Enhancing electric vehicle sustainability through battery life optimal charging. Transp. Res. Part B Methodol. **112**, 1–18 (2018)
6. Perez, H.E., Hu, X., Dey, S., Moura, S.J.: Optimal charging of Li-Ion batteries with coupled electro-thermal-aging dynamics. IEEE Trans. Veh. Technol. **66**, 7761–7770 (2017)
7. Su, W., Chow, M.-Y.: Computational intelligence-based energy management for a large-scale PHEV/PEV enabled municipal parking deck. Appl. Energy **96**, 171–182 (2012)
8. Sundstrom, O., Binding, C.: Flexible charging optimization for electric vehicles considering distribution grid constraints. IEEE Trans. Smart Grid. **3**, 26–37 (2012)
9. Alonso, M., Amaris, H., Germain, J., Galan, J.: Optimal charging scheduling of electric vehicles in smart grids by heuristic algorithms. Energies **7**, 2449–2475 (2014)
10. IEC 61851-An. In: International Standard for Electric Vehicle Conductive Charging Systems. (2018)
11. Houbbadi, A., Trigui, R., Pelissier, S., Redondo-Iglesias, E., Bouton, T.: Optimal scheduling to manage an electric bus fleet overnight charging. Energies **12**, 2727 (2019)
12. Rogge, M., Wollny, S., Sauer, D.: Fast charging battery buses for the electrification of Urban public transport—a feasibility study focusing on charging infrastructure and energy storage requirements. Energies **8**, 4587–4606 (2015)
13. Liu, Z., Yan, Y., Qu, X., Zhang, Y.: Bus stop-skipping scheme with random travel time. Transp. Res. Part C **35**, 46–56 (2013)

14. Lee, J., Lee, B.J., Park, G. L., and Kim, Y. C. Web service-based tour-and-charging scheduler framework for rent-a-car systems employing electric vehicles, Int. J. Control Autom., **6**(4), (2013)
15. Lan, T., Hu. J., Kang, Q., Si, C., Wang, L., Wu, Q.: Optimal control of an electric vehicles charging schedule under electricity markets, Neural Comput. Appl., **23**, 7–8 (2013)
16. Aabrandt, A., Andersen, P., Pedersen, A., You, S., Poulsen B., O'Connell, N., Ostergaard, J.: Prediction and optimization methods for electric vehicle charging schedules in the edison project. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT) (2012)
17. He, Y., Venkatesh, B., Guan L.: Optimal scheduling for charging and discharging of electric vehicles, IEEE Trans. Smart GRID, **3**(3) (2012)
18. Wang, S., Qu, X.: Station choice for Australian commuter rail lines: equilibrium and optimal fare design. Eur. J. Oper. Res. **258**(1), 144–154 (2017)
19. Kezunovic, M., Waller, S.T. and Damnjanovic, I.: Framework for studying emerging policy issues associated with phevs in managing coupled power and transportation systems. In: 2010 IEEE Green Technologies Conference, pp. 1–8 (2010)
20. Galus, M.D., Andersson, G.: Demand management of grid connected plug-in hybrid electric vehicles (PHEV). In 2008 IEEE energy 2030 conference pp. 1–8 (2008)
21. Lindgren, L.: Full electrification of Lund city bus traffic—a simulation study. Lund University, Lund, Sweden (2015)
22. Kunith, A., Mendelevitch, R., Goehlich, D.: Electrification of a city bus network–an optimization model for cost-effective placing of charging infrastructure and battery sizing of fast charging electric bus systems. Int. J. Sustain. Transp. **11**, 707–720 (2017)
23. Wang, S., Zhang, W., Qu, X.: Trial-and-error train fare design scheme for addressing boarding/alighting congestion at CBD stations. Transp. Res. Part B **118**, 318–335 (2018)
24. Meng, Q., Qu, X.: Bus dwell time estimation at a bus bay: A probabilistic approach. Transp. Res. Part C, **36,** 61–71 (2013)
25. Emre, M., Vermaat, P., Naberezhnykh, D., Damausuis, Y., Theodoropoulos, T., Cirimele, V., Doni, A.: Review of existing power transfer solutions, FABRIC (2014)
26. Choi, S.Y., Gu, B.W., Jeong, S.Y., Rim, C.T.: Advances in wireless power transfer systems for roadway powered electric vehicles. IEEE J. Emerg. Sel. Top. Power Electron. **3**(1), 18–36 (2015)
27. Onar, O.C., Miller, J.M., Campbell, S.L., Coomer, C., White, C.P., Seiber, L.E.: A novel wireless power transfer system for in-motion EV/PHEV charging. In: Proceedings of the IEEE Applied Power Electronics Conference and Exposition (APEC), Mar 17–21, Long Beach (2013)

Vehicle Scheduling Model for an Electric Bus Line

Jinhua Ji [iD]**, Yiming Bie** [iD]**, and Bin Shen** [iD]

Abstract The promotion of electric buses is of great significance for reducing vehicle emission, decreasing operation costs of transit corporations and workloads of bus drivers. However, the adoption of electric buses is constrained by their limited driving range. To guarantee the regular level of service, electric buses need to get recharged during daily operating hours. Electric bus battery life is highly correlated to charging modes. In this study, we proposed a mixed charging strategy with the setup of lower and upper limits of battery state of charge (SOC). A bi-level optimization model for electric bus scheduling was developed considering bus fleet size, variance of travel times of all buses and their idling times. The lower-level model is to minimize the variance of travel times and to maximize the average idling times of all buses. The upper-level model is to minimize the extra economic cost resulting from the bus fleet expansion. A case study is conducted to assess the proposed optimization model with a real electric bus route. The results show that the proposed model is capable of maintaining bus battery SOC within the reasonable range.

Keywords Electric bus · Static scheduling · Mixed charging strategy · Battery state of charge · Bi-level optimization model

J. Ji · Y. Bie (✉)
Jilin University, Changchun, Jilin Province 130022, China
e-mail: yimingbie@126.com

J. Ji
e-mail: jinhua_ji@126.com

B. Shen
School of Civil Engineering and Transportation, South China University of Technology,
Guangzhou 510641, China
e-mail: cybinshen@163.com

1 Introduction

In terms of advantages such as low noise level, zero emission and high driving stability, the promotion of electric buses is of great significance for reducing vehicle emissions in urban areas, decreasing operation costs of transit corporations and workloads of bus drivers. Nowadays, the number of electric buses deployed in the transit operation in China has accounted for 90% of the global market [1]. Besides, a number of plans have been proposed for adoptions of electric buses in developed countries, such as the TIGER program in the USA and the Green Bus Fund Program in the UK [2].

Most research efforts of bus scheduling focus on scheduling of conventional fuel buses [3, 4]. Many literatures related to bus operation service provide reference and basis for this paper. Qu et al. established an electric vehicle following model to reduce the electric energy consumption caused by traffic oscillation caused by drivers, based on reinforcement learning [5]. Wang et al. derived a solution algorithm for obtaining a user equilibrium and proposed an analytical solution approach [6]. Meng et al. proposed a regenerative stochastic process to formulate bus dwell time [7], and Liu et al. proposed optimization models for optimizing these two strategies: stop skipping and deadheading [8]. Wang et al. developed a bi-objective model and proved the existence of Pareto optimal solution mathematically without explicit demand function [9].

Comparing to fuel buses, the adoption of electric buses is constrained by their limited driving range. Therefore, some researchers recently took electric bus features into consideration, proposing new scheduling methods or conducting evaluations on influences of electric buses [10]. For example, Li examined key technical specifications that were critical to the operations of electric bus systems, in particular the operational distance and charging time [2]. Valenti et al. evaluated the changes in energy demand and resulting air pollutant emissions from the introduction of battery electric buses into the public transport operating in Rome [11]. Xylia et al. developed a dynamic optimization model for establishing charging infrastructure for electric buses by taking Stockholm, Sweden, as a case study [12]. Jiang et al. developed a neighborhood search-based heuristic considering charging and dispatching policies for scheduling the charging electric bus on city routes [13].

Electric bus battery life is highly correlated to charging modes. The most common charging mode would dramatically decrease the battery life which is to charge electric buses when batteries almost get drained and get them fully recharged every time, so as to increase driving ranges. One way to increase electric bus battery life is to maintain the state of charge (SOC) around 50% as long as possible. From the aspect of prolonging the battery life, we proposed a mixed charging strategy with the setup of lower and upper limits of battery SOC and designed a bi-level optimization model with minimizing variance of travel times during daily operating hours and maximizing the average idling time of all buses and the bus fleet to generate the optimal scheduling scheme.

2   Model Formulation

2.1   Problem Description and Formulation

The vehicle scheduling is to determine which trips are served by each bus with the given number of total trips N required by a bus route. It is assumed that there are M stops in each direction along an electric bus line, inbound trips start from stop 1 and terminate at stop M and outbound trips is opposite. Let m be the serial number of stops and $1 \leq m \leq M$ (Fig. 1). Operating hours in one day are divided into I time intervals, and i is the serial number of time intervals and $1 \leq i \leq I$.

To prolong the battery life, we set up upper and lower limits for battery SOC, which are denoted by λ_1 and λ_2. Battery SOC needs to be guaranteed within the safety threshold during scheduling. The mixed charging strategy proposed in this paper contains two parts: (i) Buses get recharged when they arrive at stop 1 and stop M. $\widehat{T}_1^i$ and $\widehat{T}_M^i$ are the charging time assigned at stop 1 and stop M in time interval i. The durations of charging time are variant during different time intervals at stop 1 and stop M. (ii) Buses get recharged when they are under the idle state. Idling time is defined as the time duration between the end time of the current trip and the start time of the next trip served by the same electric bus. Some buses are under the idle state especially during off-peak hours.

Let $x_{n_{\text{up}}}^k$ denote the relation between electric buses k. It is equal to one if bus k serves the trip n_{up}, and 0 otherwise. Likewise, $y_{n_{\text{down}}}^k$ is the relation between electric bus k and outbound trip n_{down}. $\widehat{T}_{1,M}^i$ and $\widehat{T}_{M,1}^i$ represent the average travel times of inbound and outbound trips, respectively. They can be obtained by historical data. Let t_c^k denote the departure time of the cth trip served by bus k, and t_d^k denote the departure time of the dth trip served by bus k. The idling time of bus k at stop 1 before the cth inbound trip and at stop M before the dth outbound trip are:

$$T_c(k) = t_c^k - \left(t_d^k + \widehat{T}_{M,1}^i \right) - \widehat{T}_1^i \quad (c \neq 1) \tag{1}$$

$$T_d(k) = t_d^k - \left(t_{c-1}^k + \widehat{T}_{1,M}^i \right) - \widehat{T}_M^i \quad (d \neq 1) \tag{2}$$
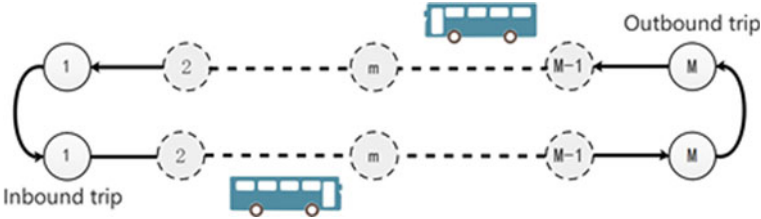
Fig. 1   Schematic diagram of an electric bus line

Equations ([1](#)) and ([2](#)) are required to satisfy the constraint $t_{c-1}^k \leq t_d^k \leq t_c^k \leq t_{d+1}^k$.

2.2 Battery State Function

Based on the proposed charging strategy, we set up the upper limits for charging time at stop 1 and stop M. During the same time interval, the charging time limits are identical and changed for different time intervals.

The variation of battery SOC during charging is represented by $h(t)$, where C_N is rated capacity of the battery, kW h. I is the charging current, A. U is the charging voltage, V. η denotes coulombic efficiency, of which the default value is 1. t is the charging time, m.

$$h(t) = \frac{\eta U I}{60\, C_N} t \qquad (3)$$

During discharging, the energy consumption rate increases gradually with the reduction of battery remaining energy. The time-variant curve of battery energy is a decreasing concave function in the first quadrant. Therefore, the variation of battery SOC during discharging is jointly determined by the initial SOC and the travel time:

$$\varphi(\text{SOC}_0, t) = \text{SOC}_0 + a_1 t^3 + a_2 t^2 + a_3 t \qquad (4)$$

Function φ can be fitted with real electric bus operating data.

The varying of the battery SOC of electric bus k between its departure from stop 1 at t_c and its departure from stop M at t_d can be described by the following Eq. ([5](#)). Similarly, the varying of the battery SOC between the departure from stop M at t_d and its departure from stop 1 at t_{c+1} can be described by the following Eq. ([6](#)).

$$\text{SOC}_M^k(t_d) = \begin{cases} \text{SOC}_1^k(t_c) - \varphi\left[\text{SOC}_1^k(t_c), \widehat{T}_{1,M}^i\right] + h(T_d(k)) & \text{if } T_d(k) \leq T_d'(k) \\ \lambda_1 & \text{if } T_d(k) > T_d'(k) \end{cases} \qquad (5)$$

$$\text{SOC}_1^k(t_{c+1}) = \begin{cases} \text{SOC}_M^k(t_d) - \varphi\left[\text{SOC}_M^k(t_d), \widehat{T}_{M,1}^i\right] + h(T_c(k)) & \text{if } T_c(k) \leq T_c'(k) \\ \lambda_1 & \text{if } T_c(k) > T_c'(k) \end{cases} \qquad (6)$$

where $T_d'(k)$ is the maximum available charging time for electric bus k at stop M and $T_c'(k)$ is the maximum available charging time for electric bus k at stop 1.

2.3 Optimization Model Formulation

During the bus operation, extending the charging time and idling time at stop 1 and stop M is beneficial to prolonging battery life, which however would reduce the number of buses available for deploying. Consequently, bus companies have to purchase new buses to satisfy operation demands, which would increase the operation costs. Therefore, we proposed a bi-level optimization model to optimize the static scheduling scheme for operating hours in a day.

Upper-level Optimization Model The serial numbers of buses dispatched in time interval i are denoted by $R(i)$ and can be obtained based on the bus service schedule.

The upper-level model is to minimize the extra cost due to the bus fleet expansion, satisfying the constraints of passenger demands in each time interval. The model can be formulated as follows:

$$\min W = \overline{D}(K - K_0) \tag{7}$$

$$\text{s.t.} \quad K \geq \max\left(k^1, k^2, k^3, \ldots, k^I\right) \tag{8}$$

$$|R(i)| \geq \text{ceil}\left(\frac{\widehat{T}_1^i + \widehat{T}_M^i + \widehat{T}_{1,M}^i + \widehat{T}_{M,1}^i}{H^i}\right) \tag{9}$$

In Eq. (7), W is the extra cost due to the bus fleet expansion and $\overline{D}$ is the cost of purchasing an electric bus. Equations (8) and (9) represent the constraints of required bus fleet size for operation schedule. $|R(i)|$ is the number of elements containing in the set.

Lower-level Optimization Model Total travel time of electric bus k in a day can be calculated as follows:

$$T_1(k) = \sum_{n_{\text{up}}=1}^{N_{\text{up}}} \widehat{T}_{1,M}^i x_{n_{\text{up}}}^k + \sum_{n_{\text{down}}=1}^{N_{\text{down}}} \widehat{T}_{M,1}^i y_{n_{\text{down}}}^k \tag{10}$$

The lower-level optimization model which is to minimize the variance of travel times and to maximize the average idling time of all buses is given by:

$$\min Z = \alpha_1 \frac{1}{K} \sum_{k=1}^{K} \left[T_1(k) - \frac{1}{K} \sum_{k=1}^{K} T_1(k) \right]^2$$

$$- \alpha_2 \frac{1}{K} \sum_{k=1}^{K} \left[\sum_{c=2}^{C_{\text{up}}(k)} T_c(k) + \sum_{d=2}^{C_{\text{down}}(k)} T_d(k) \right] \tag{11}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} x_{n_{\text{up}}}^{k} = 1 \tag{12}$$

$$\sum_{k=1}^{K} y_{n_{\text{down}}}^{k} = 1 \tag{13}$$

$$t_{c-1}^{k} + \widehat{T}_{1,M}^{i} + \widehat{T}_{M}^{i} \leq t_{d}^{k} \leq t_{c}^{k} - \widehat{T}_{M,1}^{i} - \widehat{T}_{1}^{i} \quad (c \neq 1, d \neq 1) \tag{14}$$

$$t_{d}^{k} + \widehat{T}_{M,1}^{i} + \widehat{T}_{1}^{i} \leq t_{c}^{k} \leq t_{d+1}^{k} - \widehat{T}_{1,M}^{i} - \widehat{T}_{M}^{i} \quad (c \neq 1, d \neq 1) \tag{15}$$

$$\text{SOC}_{1}^{k}(t_c) \geq \lambda_2 + \varphi\left[\text{SOC}_{1}^{k}(t_c), \widehat{T}_{1,M}^{i}\right] \tag{16}$$

$$\text{SOC}_{M}^{k}(t_d) \geq \lambda_2 + \varphi\left[\text{SOC}_{M}^{k}(t_d), \widehat{T}_{M,1}^{i}\right] \tag{17}$$

Equation (11) gives the objective function of the electric bus scheduling optimization model; α_1 and α_2 are weighted parameters, $0 \leq \alpha_1 \leq 1$, $0 \leq \alpha_2 \leq 1$ and $\alpha_1 + \alpha_2 = 1$. Equations (12) and (13) refer to that each inbound or outbound trip can only be served by one electric bus. Equations (14) and (15) are constraints of continuous traveling time of bus k. The battery SOC of electric buses should satisfy Eqs. (16) and (17) every time when they depart from stop 1 and stop M, respectively.

2.4 Solution Algorithm

The optimization model proposed is a bi-level multi-objective nonlinear optimization model, and solving this is a NP-hard problem. Genetic algorithm (GA) is capable of conducting global optimal search in the complex solution domain and has strong robustness when solving this kind of problem.

We use binary encoding in this algorithm. Bus fleet size K of the upper-level model is selected as the initial population. Solving the lower-level problem using GA and the feasible value of fleet size can be obtained from the upper level optimization population. Then, with function Z from the Eq. (11) as the fitness function, evaluations are conducted and followed by operations such as selection, crossover and mutation. Optimal charging times and scheduling schemes of different fleet sizes obtained from the previous steps are recorded and sent back to the upper level. Compare and evaluate the upper fitness function values of all feasible individuals. We put the fleet sizes into the objective function W in the upper-level problem and evaluate the fitness function values of extra operation costs caused by bus fleet expansion. Individuals with higher values have higher probability to breed a new generation. The fitness function values for infeasible individuals are set to be zero.

3 Numerical Example

3.1 Bus Route Description

We take the 107 electric bus lines in Meihekou City as an example to assess the proposed scheduling model. This bus line is 15 km long with 17 bus stops. The electric bus model is E8PLUS with 180 km driving range. The length of this model is 8.5 m and the weight is 7040 kg. The operation time of inbound trips is 06:20–20:00 and that of outbound trips is 06:35–20:35. The operation hours of this route are divided into four periods. The departure headway H^i of a bus line during time interval i is determined by the maximum sectional passenger flow and was 10, 20, 15 and 30 min, respectively.

3.2 Analysis of Results

Corresponding charging times and average travel time are illustrated in Table 1 and bus departure schedules are illustrated in Table 2. The bi-level optimization problem was solved with the genetic algorithm proposed in Sect. 2, and the solution indicates that the optimal fleet size K is 13.

Total travel times and total idling times of 13 electric buses are shown in Fig. 2. Daily total travel times of the first four vehicles are slightly longer than other buses for each of them serves one more trip. Their average travel time is 7.6 h in a day, and the variance of their travel times is 0.01. The average travel time of other nine electric buses is 6.7 h, and the variance of their travel times is 0.02. The average idling time for all buses is 5.9 h, which indicates that all buses have plenty of idling times to get recharged in time, avoiding excessive discharge and its impact on battery life.

Variations of battery SOC of all electric buses in a day are depicted in Fig. 3. The battery SOC fluctuates within (65%, 85%); none of which has been below the lower limit of the safety threshold. Besides, we noticed that during morning peak hours, the SOC continuously drops. This is because buses continuously serve to satisfy passenger demands. In our charging strategy, charging times are assigned to buses at

Table 1 Average travel times in different time periods and the optimal charging times at stop 1 and stop 17 under different time periods

Time periods	Charging times (min)		Average travel time (min)	
	Stop 1	Stop 17	Inbound trip	Outbound trip
6:20–10:00	10	8	57	55
10:00–15:00	15	13	46	43
15:00–18:00	12	10	51	54
18:00–20:30	18	17	38	40

Table 2 Optimal bus departure times for route 107

No	Electric bus departure time								
	Stop 1	Stop 17	Stop 1	Stop 17	Stop 1	Stop 17	Stop 1	Stop 17	Stop 1
1	6:20	7:25	8:30	9:35	11:25	13:30	15:32	17:07	19:39
2	6:35	7:40	8:45	9:50	11:50	14:05	15:52	17:32	20:14
3	6:30	7:35	8:40	9:45	11:45	13:50	15:47	17:22	20:09
4	6:45	7:50	8:55	10:05	12:10	14:25	16:07	17:47	20:44
5	6:40	7:45	8:50	9:55	12:05	14:10	16:02	17:37	–
6	6:55	8:00	9:05	10:25	12:30	14:45	16:22	18:09	–
7	6:50	7:55	9:00	10:10	12:25	14:30	16:17	17:52	–
8	7:05	8:10	9:15	10:45	12:50	15:02	16:37	18:39	–
9	7:00	8:05	9:10	10:30	12:45	14:50	16:32	18:14	–
10	7:15	8:20	9:25	11:05	13:10	15:17	16:52	19:09	–
11	7:10	8:15	9:20	10:50	13:05	15:07	16:47	18:44	–
12	7:20	8:25	9:30	11:10	13:25	15:22	17:02	19:14	–
13	7:30	8:35	9:40	11:30	13:45	15:37	17:17	19:44	–
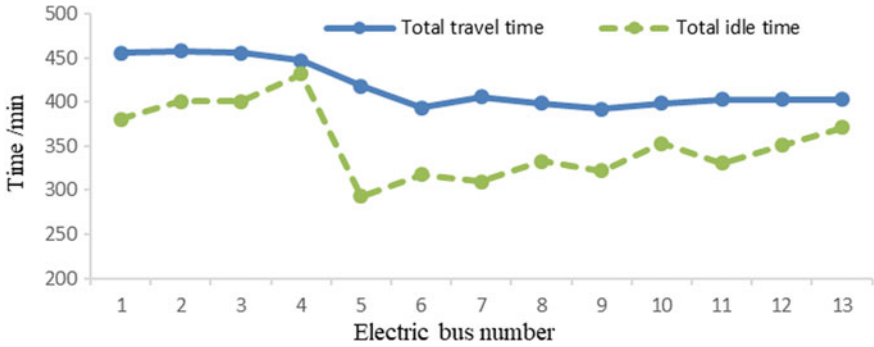
Fig. 2 Daily total travel times and idle times of all electric buses
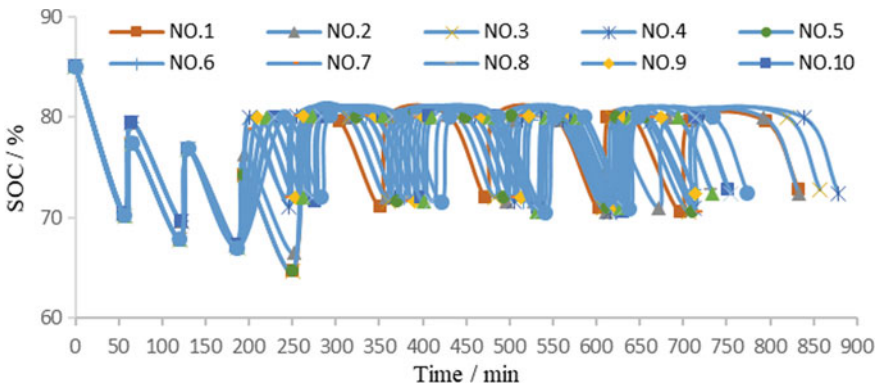
Fig. 3 Variations of the battery SOC with time

both start and terminal stops, which effectively extend the driving ranges of electric buses.

4 Conclusion

In terms of prolonging the battery life, we proposed a mixed charging strategy in this paper with the setup of lower and upper limits of battery SOC. A bi-level optimization model for electric bus scheduling was developed considering bus fleet size, variance of travel times of all buses and their idling times. We carried out the assessment of the proposed optimization model with a real-world example of the 107 bus line in Meihekou City. Some conclusions were drawn based on the results as follows:

(i) Reasonable assignment of trips to each bus is of great significance to increasing average bus idling time and reducing the cost of fleet expansion. Buses getting

recharged during idling time do not affect the normal scheduling of a route. Instead, it helps to maintain battery SOC within the ideal range.

(ii) Based on the analysis and results of the example, the proposed bi-level optimization model is capable of maintaining bus battery SOC within the range (65%, 85%), which has an important impact on extending battery life.

How the battery rated capacity drop influences bus scheduling will be considered in the future research. When the battery rated capacity decreases to a certain extent, the current fleet size cannot satisfy the requirements of keeping normal operation of a bus route. In terms of economic costs, the influences of different strategies, such as adjusting the upper and lower limits of SOC safety threshold, changing batteries or purchasing new electric buses, should be evaluated to make the optimal bus scheduling scheme.

References

1. Tang, X., Lin, X., He, F.: Robust scheduling strategies of electric buses under stochastic traffic conditions. Transp. Res. Part C Emerg. Technol. **105**, 163–182 (2019)
2. Li, J.Q.: Battery-electric transit bus developments and operations: a review. Int. J. Sustain. Transport. **10**(3), 157–169 (2016)
3. Ceder, A.: Public-transport vehicle scheduling with multi vehicle type. Transp. Res. Part C Emerg. Technol. **19**(3), 485–497 (2011)
4. Bie, Y., Xiong, X., Yan, Y., Qu, X.: Dynamic headway control for high-frequency bus line based on speed guidance and intersection signal adjustment. Comput. Aided Civil Infrastruct. Eng. **35**, 4–25 (2020)
5. Qu, X., Yu, Y., Zhou, M., Lin, C.T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach. Appl. Energy **257**, 114030 (2020)
6. Wang, S., Qu, X.: Station choice for Australian commuter rail lines: equilibrium and optimal fare design. Eur. J. Oper. Res. **258**(1), 144–154 (2017)
7. Meng, Q., Qu, X.: Bus dwell time estimation at a bus bay: a probabilistic approach. Transp. Res. Part C **36**, 61–71 (2013)
8. Liu, Z., Yan, Y., Qu, X., Zhang, Y.: Bus stop-skipping scheme with random travel time. Transp. Res. Part C **35**, 46–56 (2013)
9. Wang, S., Zhang, W., Qu, X.: Trial-and-error train fare design scheme for addressing boarding/alighting congestion at CBD stations. Transp. Res. Part B **118**, 318–335 (2018)
10. Vepsäläinen, J., Baldi, F., Lajunen, A., Kivekäs, K., Tammi, K.: Cost-benefit analysis of electric bus fleet with various operation intervals. In: 2018 21st International Conference on Intelligent Transportation Systems, Maui, HI, USA, pp. 1522–1527 (2018)
11. Valenti, G., Liberto, C., Lelli, M., Ferrara, M., Nigro, M., Villante, C.: The impact of battery electric buses in public transport. In: 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe, Milan, Italy, pp. 1–5 (2017)

12. Xylia, M., Leduc, S., Patrizio, P., Kraxner, F., Silveira, S.: Locating charging infrastructure for electric buses in Stockholm. Transp. Res. Part C Emerg. Technol. **78**, 183–200 (2017)
13. Jiang, M., Zhang, Y., Zhang, C., Zhang, K., Zhang, G., Zhao, Z.: Operation and scheduling of pure electric buses under regular charging mode. In: 2018 21st International Conference on Intelligent Transportation Systems, Maui, HI, USA, pp. 1894–1899 (2018)

How to Model the Influence of In-vehicle Crowding on Travel Behavior: A Comparison Among Moderation, Independent Variable and Interaction

Kun Gao , **Jieyu Fan, and Ziling Zeng**

Abstract Accurate modeling of travel choice behavior is crucial for effective transport demand forecasting, management and planning. This study tries to shed light on the appropriate modeling approach concerning the influences of in-vehicle crowding on mode choice behavior in the multimodal network. Stated preference surveys covering four commuting transport modes and four influencing factors are conducted to collect empirical behavior data. Three modeling methods, treating the in-vehicle crowding as a moderator of perceived travel time, as an independent variable and by incorporating interaction effect, are empirically compared. The result indicates that there is a bidirectional interaction between travel time and in-vehicle crowding. The influence of in-vehicle crowding increases with increasing travel time and vice versa. Considering crowding as an independent variable and taking the effects of travel time on the perception of in-vehicle crowding are the best ways to depict the overall influences of in-vehicle crowding. The sensitivity analysis shows that increasing the cost of using car is comparatively effective for reducing car usage. Shortening the travel time of public transit and improving service quality such as travel time reliability and in-vehicle crowding are more useful in attracting car users as compared to reduction in the cost of public transit. The results provide insights into travelers' behavior in the multimodal network and could support scientific transport management and planning.

Keywords Travel behavior · Sustainable transport system · In-vehicle crowding · Multimodal network

K. Gao (✉) · Z. Zeng
Department of Architecture and Civil Engineering, Chalmers University of Technology, 412 96
Goteborg, Sweden
e-mail: gkun@chalmers.se

J. Fan
College of Transportation Engineering, Tongji University, Shanghai, China

1 Introduction

Transportation Operations Annual Report of Shanghai 2018 [1] reported that the average travel speed during peak hours in arterial roads of the central city in 2018 decreased by 6–10% compared to that in 2017. Regardless of increasing traffic congestion, car remains the most attractive mode for commuting in Chinese metropolis like Shanghai due to its flexibility and comfort in contrast to overcrowding public transit in peak hours. Promoting the share rate of public transit (PT) is an inevitable trend for large metropolises with huge populations to shape a more sustainable transportation system [2, 3]. However, several taken measures for improving the share rate of PT in Shanghai (e.g., establishment of new transit facilities, discount for using park and ride and giving road priority to transits) failed to attract car users [4]. Transport engineers and practitioners are eager to figure out the effective management strategy to adjust the mode split structure. This requires accurate understandings of travelers' mode choice behavior in the current multimodal network.

Travel behavior in the multimodal network is complicated since a variety of modes are available for the travelers including metro, bus, walking, cycling, car and combined modes (e.g., P&R), and a plenty of attributes are potentially considered by the traveler containing cost, journey time, waiting time, reliability, crowding, transfer time, etc. [5]. Especially, with the upturn living standards, travelers seem to attach more and more weights to the service qualities and comfortable features like travel time reliability (TTR) and in-vehicle crowding levels rather than travel time or cost [6–8]. The in-vehicle crowding denotes a high density of passage in the carriage that produces much unpleasure both physically and mentally. The effects of in-vehicle crowding on travel behavior are not easily measurable and have complex mechanisms. For instance, travelers' aversion to in-vehicle overcrowding is not only simply due to physical discomfort but also ascribed to the fact that overcrowding reduces travelers' tolerance for other level-of-service factors [9, 10].

Improving the crowding levels of public transit is recognized to be one of the key components for PT service quality improvement [11–14], design and planning of PT system [14, 15]. Overcrowding in PT discouraged travelers' willingness to use PT for commuting and increased travelers' perceived negative effects of time in the carriage, walking and waiting time [11, 16]. Commonly, in-vehicle crowding is quantitatively measured by two approaches: One is load factor [17] defining as the amount of passengers in vehicle divided by the number of seats; another measurement is standing density [6] with a unit of passengers per square meter (person/m^2) indicating the amount of standing passengers divided by the area in vehicle. The load factor is easily computed based on the number of onboard passengers but is somehow inefficient in accurately describing the degree of crowding suffered by standing passengers. In this aspect, standing density performs better in capturing situations with many standing passengers [18].

Many studies have demonstrated that in-vehicle crowding significantly influences travelers' mode choice behavior. There are different perspectives and models for considering the influencing of in-vehicle crowding on travel behavior. Some scholars

deemed that in-vehicle crowding mainly influenced the perceived value of travel time (VOT) in the carriage, and the effects of crowding could be expressed as an increasing perceptional weight on in-vehicle time [19]. Shao et al. [20] evaluated the VOT in different crowding levels and demonstrated that considering the changes of VOT with different crowding levels could nicely capture the effects of in-vehicle crowding. This approach actually regards the in-vehicle crowding as a moderation factor of perceptions of travel time. The time multiplier expressed by the ratio of the marginal utility of travel time under crowded conditions to that under the uncrowded condition is generally used to model the moderation effect of in-vehicle crowding. However, others believe that in-vehicle crowding should be treated as an independent attribute in model specifications to reflect its influences, rather than a moderation effect. Using an independent variable to model the effects of in-vehicle crowding, Hirsch and Thompson [21] conducted qualitative and quantitative studies to gain insight into the passengers' perceptions and tolerance of railway crowding and analyzed its influences on travel behavior. Wardman and Murphy [22] estimated the values of crowding levels in train. The third approach is considering the in-vehicle crowding as a variable in model and simultaneously adds an interaction term to depict its influences on perceived travel time. It is still inconclusive that which way is better to comprehensively reflect and measure the influences of crowding in behavioral modeling. Further explorations to identify the appropriate method are needed to avoid inaccurate consideration of crowding levels in forecasting models.

This study tries to fill the above gap and endeavors to investigate the appropriate modeling approach concerning the influences of in-vehicle crowding based on empirical mode choice behavior data in the multimodal network. Stated preference surveys containing four commuting transport modes and four influencing factors are designed using the efficient design method to collect reliable travel behavior data. Three modeling approaches, namely regarding the in-vehicle crowding as a moderator of perceived travel time, as an independent variable and by incorporating interaction effect, are empirically compared. Moreover, the nonlinear effects of in-vehicle crowding on perceived travel time are also explored for more accurate modeling. Sensitivity analysis is conducted to investigate how the share rates of different transport modes respond to changes in level-of-service variables and thus provide implications for effective transport management to shift car users to PT.

This paper is structured as follows: Sect. 2 gives a detailed description about survey design, data collections and the model specifications. The results and analysis are presented in Sect. 3. Section 4 discusses the main findings and presents conclusions.
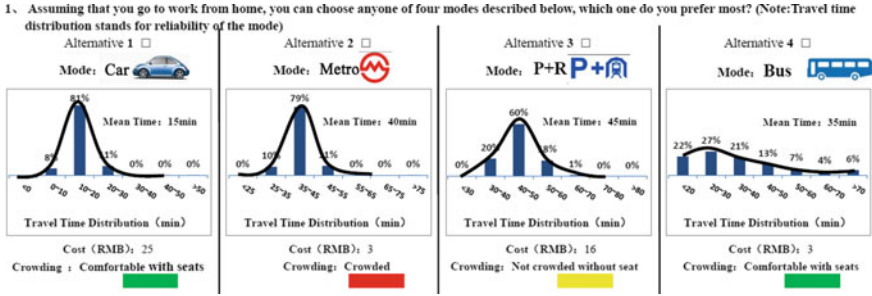
Fig. 1 An example of SP choice scenarios [23]

2 Methodology

2.1 Survey Design

A two-stage survey was carried out to investigate travelers' mode choice in multimodal networks. The first-stage survey collected the general trip information of commuters in Shanghai to identify the most decisive attributes influencing travelers' mode choice behavior, which was then used for the design of the second-stage survey. In the second stage, an experimental survey questionnaire was designed to investigate how travelers value the attributes of each mode quantitatively. Four commonly used modes for commuting in Shanghai are incorporated including car, metro, park and ride (P&R) and bus. Decisive attributes including cost, travel time, travel time reliability and in-vehicle crowding are considered. The detailed information about the survey design is available in Gao et al. [23]. It is worthy of mentioning that three levels of crowding were set, and the value of standing density of each crowding level was according to the classification principles of crowding levels in the relevant literature [20, 24]. The travel time reliability is measured by standard deviation of travel time. One example of the scenarios is illustrated in Fig. 1. With the assistance of traffic police department, we conducted surveys in two working halls of the Bureau of Vehicle Management in Shanghai and public transit hubs. Face-to-face surveys were carried out, and respondents were requested to read and understand questions carefully with the assistance of investigators. Finally, 386 (2318 observations) effective questionnaires were collected. Attributes of all respondents are summarized in Table 1.

2.2 Model Specifications and Estimation

Discrete choice technique and expected utility theory are applied to model how travelers value attributes of each mode and make the travel decisions. Five utility

Table 1 Summary of effective respondents [5]

Age	N	%	Education	N	%	Income (CNY)	N	%
<30	158	41	1	103	27	<3000	19	5
30–40	153	40	2	176	46	3000–6000	105	27
40–50	50	13	3	81	21	6000–10,000	139	36
>50	13	3	4	9	2	10,000–20,000	77	20
						>20,000	26	7
Skipped	12	3		17	4		20	5
Total	386	100		386	100		386	100
Gender	N	%	Car owner	N	%	License type	N	%
Male	245	63	Yes	324	84	Shanghai	253	78
Female	141	37	No	62	16	Non-Shanghai	71	22
Total	386	100		386	100		324	100

Note Education 1: Lower than undergraduate; 2: Undergraduate; 3: Master; 4: Doctor. 1 CNY = 0.14 dollar

specifications are proposed as shown in Eq. (1)–(5) to examine different modeling approaches concerning the influences of in-vehicle crowding. Model 1 treats the in-vehicle crowding as a moderation factor of perceived travel time [19] but uses linear functions. Model 2 follows the same modeling method in Model 1 and further applies a nonlinear function according to Shao et al. [20]. Model 3 treats in-vehicle crowding as an independent variable in the utility function. Model 4 treats in-vehicle crowding as a decisive variable and simultaneously considers the interaction effects of in-vehicle crowding and travel time. The potential nonlinearity of the effect of in-vehicle crowding on perceived travel time is considered as well. Model 5 uses a similar modeling approach as Model 4 but considers the nonlinear effects of travel time on perceived in-vehicle crowding.

$$\text{Model 1:} \quad U_i = ASC_i + \beta_c * C_i + (\beta_{tt} + \beta_{tcr} * \text{crowd}_i) * TT_i + \beta_{re} * TTR_i \quad (1)$$

$$\text{Model 2:} \quad U_i = ASC_i + \beta_c * C_i + (\beta_{tt} + \alpha * \text{crowd}_i^\lambda) * TT_i + \beta_{re} * TTR_i \quad (2)$$

$$\text{Model 3:} \quad U_i = ASC_i + \beta_c * C_i + \beta_{tt} * TT_i + \beta_{re} * TTR_i + \beta_{cr} * \text{crowd}_i \quad (3)$$

$$\text{Model 4:} \quad U_i = ASC_i + \beta_c * C_i + (\beta_{tt} + \alpha * \text{crowd}_i^\lambda) * TT_i$$
$$+ \beta_{re} * TTR_i + \beta_{cr} * \text{crowd}_i \quad (4)$$

$$\text{Model 5:} \quad U_i = ASC_i + \beta_c * C_i + \beta_{tt} * TT_i + \beta_{re} * TTR_i$$
$$+ (\beta_{cr} + \alpha * TT_i^\lambda) * \text{crowd}_i \quad (5)$$

where i is the index of transport mode. U_i is the determined utility of mode i. ASC_i denotes the constant for mode i. C_i, TT_i, TTR_i and crowd_i represent cost, regular travel time, travel time reliability and crowding levels, respectively. Others are coefficients to be estimated.

Random parameter logit (RPL) model is adopted for the model estimation since it is capable of capturing travelers' heterogeneities in preference. Accounting for that a respondent actually made a sequence of scenarios in one questionnaire, panel data process is performed to consider panel effects in estimation. An error component model is performed to capture the possible correlation between modes. However, it turned out that error components were not significant. Pythonbiogeme is used to implement the estimation process.

3 Result and Analysis

The results are shown in Table 2. It can be observed that the coefficient of moderation effect of in-vehicle crowding on travel time in Model 1 is negative (-0.00575^{**}) and significant in the confidence level of 99%. The result demonstrates that the weight of travel time rises with increasing in-vehicle crowding levels. It is realistic that travelers are sick of travel time lasted in the crowded carriage. The result can be interpreted in another way that the longer the travel time is, the more the crowding levels are valued.

Model 2 reflects the influences of in-vehicle crowding in the utility function by considering the effects of crowding levels on increasing perceptional values of travel time. All the estimated coefficients in Model 2 are significant. The α is negative (-0.0081^{**}), implying that the travelers' perceptional values of travel time indeed have an increasing relationship with increasing crowding levels. The λ is positive (0.753^{**}) indicating that the marginal effect of in-vehicle crowding increases with the scale of crowding levels.

Model 3 regards the in-vehicle crowding as an independent variable in the utility function. All coefficients are significant in the confidence level of 95%. Akaike information criterion (AIC) as a widely used comparison method [25, 26] is adopted to compare which model is better to evaluate the influences of in-vehicle crowding

Model 3 (AIC = 4212.238) is much superior to Model 2 (AIC = 4252.802) and Model 1 (AIC = 4240.592). It reveals that in-vehicle crowding has significant influences on travelers' mode choice and should be treated as an independent variable, rather than being considered as a moderator of perceived travel time as in Model 1 and Model 2 in either linear or nonlinear ways.

Furthermore, Model 4 regards in-vehicle crowding as a variable and simultaneously considers the effects of in-vehicle crowding on the perception of travel time. The AIC of Model 4 is 4213.418 and even a little bit larger than that of Model 3, which means Model 3 outperforms Model 4. It seems that if in-vehicle crowding has already been an independent variable in the utility function, it is not necessary to consider the effects of in-vehicle crowding on perceptions of travel time anymore.

Table 2 Estimated results of five models

Parameters	Model 1		Model 2		Model 3	
	Mean	SD	Mean	SD	Mean	SD
Cost	−0.0727**(−7.08)	0.03**(6.34)	−0.0705**(−7.02)	0.0249***(5.05)	−0.0704**(−6.68)	0.0305***(4.83)
Travel time	−0.0547**(−10.23)	0.0445**(9.85)	−0.0498**(−9.25)	0.0451**(9.65)	−0.0649**(−11.07)	0.043**(8.34)
Travel time reliability	−0.0968**(−9.78)	0.104**(8.37)	−0.0961**(−9.92)	0.0949**(8.18)	−0.100**(−9.07)	0.0958**(5.48)
Crowding					−0.24**(−11.77)	0.203**(7.73)
Moderation of crowding on travel time	−0.00575**(−11.87)	0.00234*(2.07)				
α			−0.0081**(−3.55)			
λ			0.753***(4.45)			
ASC$_{bus}$	0	fixed	0	fixed	0	Fixed
ASC$_{car}$	2.26**(6.09)		2.3**(6.25)		2.11**(5.45)	
ASC$_{metro}$	1.09** (10.62)		1.09**(10.82)		1.17**(11.06)	
ASC$_{P\&R}$	1.34** (6.39)		1.35**(6.47)		1.38***(6.42)	
Final log-likelihood	−2109.296		−2115.401		−2095.119	
Rho-square	0.343		0.341		0.348	
No. of parameters	11		11		11	
AIC	4240.592		4252.802		4212.238	

Parameters	Model 4		Model 5	
	Mean	SD	Mean	SD
Cost	−0.0746**(−6.78)	0.0298**(5.62)	−0.0728**(−6.81)	0.0307**(5.83)
Travel time	−0.0588**(−8.94)	0.0424**(8.79)	−0.0602**(−10.76)	0.041**(8.42)

(continued)

Table 2 (continued)

Parameters	Model 4		Model 5	
	Mean	SD	Mean	SD
Travel time reliability	−0.0992**(−10.06)	0.0924**(7.21)	−0.0999**(−10.09)	0.0938**(7.39)
Crowding	−0.17**(−2.92)	0.202**(6.55)	−0.369**(−6.55)	0.207**(7.88)
Moderation of crowding on travel time				
α	−0.00398(−1.28)		0.226*(2.47)	
λ	0.541(0.89)		−0.571**(3.31)	
ASC_{bus}	0	Fixed	0	Fixed
ASC_{car}	2.31**(5.65)		2.2**(5.69)	
ASC_{metro}	1.18**(11.11)		1.19**(11.08)	
$ASC_{P\&R}$	1.46**(6.63)		1.29**(5.95)	
Final log-likelihood	−2093.709		−2090.106	
Rho-square	0.348		0.349	
No. of parameters	13		12	
AIC	4213.418		4204.212	

Note ** and * denote significance at 99%, 95% confidence level, respectively. Values in parentheses are the t-values in t-test
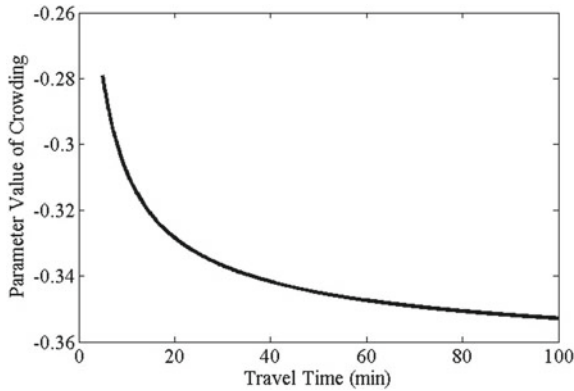
Fig. 2 Influence of travel time on the marginal effect of in-vehicle crowding

Model 5 regards in-vehicle crowding as a variable and integrates the effects of duration of travel time on crowding perception. All estimated coefficients in Model 5 are significant in the confidence level of 95%. The α is positive (0.226*) and λ is negative ($-0.571**$). It indicates that the duration of a trip indeed has significant influences on travelers' tolerance for overcrowding in the carriage. The longer the travel time is, the more weight to crowding is attached as Fig. 2 illustrates. The AIC of Model 5 is 4204.212 and is superior to Model 4 (4213.418). The results imply that taking the influences of travel time on the marginal effect of in-vehicle crowding into consideration is beneficial for improving the fitness of depicting travelers' mode choice behavior.

4 Sensitivity Analysis

The left figure in Fig. 3 demonstrates the direct elasticities of all attributes for all the four modes. Direct elasticity reflects how the share rate of a transport mode responds to changes in its own level-of-service variables. The sign of direct elasticity is negative as increases in the attributes lead to disutility. For the convenience of comparison, we adopt the absolute value of elasticity for presentation. For car, the cost has the largest average direct elasticity (1.49) followed by travel time (1.29) and travel time reliability (0.68) in sequence. The direct elasticities of cost and travel time are more than one meaning that they are elastic variables influencing the choice probability of car. This implies that increases in cost and travel time from current levels would remarkably reduce travelers' willingness to use car. For public transit, the direct elasticity of travel time is the largest with values of 2.04 and 3.03 for metro and bus, respectively. It implies that travelers' choice probability for PT is observably sensitive to the travel time of PT. The elasticity of in-vehicle crowding for bus is about 1.15, indicating that reduction in crowding level of bus is an efficient way to increase
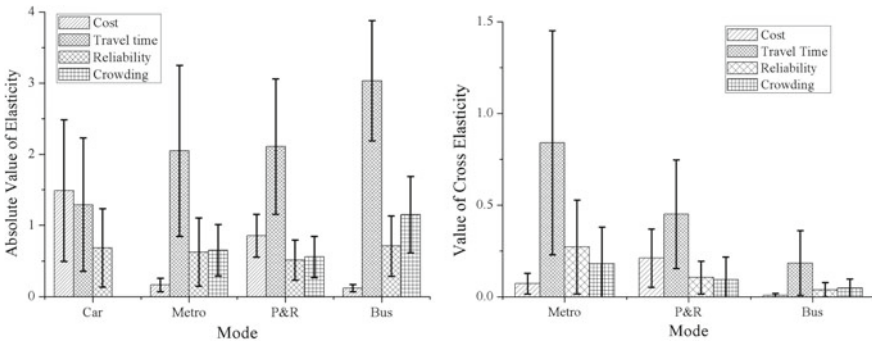
Fig. 3 Direct (left) and cross elasticity (right) of attributes

its current share rate. Travel time reliability (0.62), crowding (0.65) of metro and travel time reliability of bus (0.70) have similar elasticity. The results demonstrate that improving the reliability and crowding levels of PT do much favor to increase the attractiveness of PT. The cost of PT has extremely small elasticity meaning the respondents are not sensitive to the changes in cost of PT. For P&R, the travel time (2.1) has the largest elasticity followed by cost (0.85), in-vehicle crowding (0.56) and travel time reliability (0.51) sequentially. Reduction in travel time and cost based on the current levels is a valid measure to increase the current share rate of P&R.

The cross elasticity of attributes of metro, P&R and bus to car is illustrated in the right figure of Fig. 3. Cross elasticities of attributes of metro are larger than those of P&R on average. The cross elasticities of bus are the smallest. It indicates that improvements in attributes of metro and P&R are more efficient to attract car users shifting to PT as compared to improving the service of bus. For metro and bus, the travel time has the largest cross elasticity to car followed by travel time reliability, crowding levels and cost orderly. For P&R, travel time has the largest cross elasticity to car as well, followed by travel time reliability and in-vehicle crowding. Moreover, all the mean values of cross elasticity are less than one. It means that respondents, most of whom are chronic car users, have evident preferences for car and are not sensitive to improvements of other modes.

5   Conclusion

This study investigates the appropriate method for modeling the influences of in-vehicle crowding on mode choice based on empirical behavior data in the multimodal network. Three modeling approaches including treating the in-vehicle crowding as a moderator of perceived travel time, as an independent variable and by incorporating interaction effect are empirically compared. Sensitivity analysis is conducted to examine how the share rates of different transport modes respond to changes

in level-of-service variables and thus provide implications for effective transport management. The main findings could be summarized as follows:

- A bidirectional interaction between travel time and crowding is found. The influences of in-vehicle crowding increase with increasing travel time and vice versa. The relationship could be quantitatively measured by an interaction term with consideration of nonlinearity.
- Considering crowding as an independent variable and taking the effects of travel time on the perception of in-vehicle crowding are the best ways to depict the overall influences of in-vehicle crowding on mode choice behavior.
- The sensitivity analysis shows that increasing the cost by car could efficiently reduce car usage. Shortening the travel time of PT (metro and bus) and P&R are the most effective methods to shift car users to other modes in the contexts of Shanghai. Improvements in service quality like travel time reliability and in-vehicle crowding play crucial roles in attracting car users.

References

1. SURCTDRI, Transportation operation annual report of Shanghai (Year of 2017). 2018, Shanghai Uban and Rural Construction and Transportation Development Research Institue (SURCTDRI): Shanghai
2. Xu, C., et al.: Potential risk and its influencing factors for separated bicycle paths. Accid. Anal. Prev. **87**, 59–67 (2016)
3. Beaudoin, J., Farzin, Y.H., Lawell, C.-Y.C.L.: Public transit investment and sustainable transportation: a review of studies of transit's impact on traffic congestion and air quality. Res. Trans. Econ. (2015)
4. Gao, K., et al.: Perceptions of mode-specific travel time reliability and crowding in multimodal trip. Trans. Res. Rec. J. Transp. Res. Board, pp. 1–22 (2016)
5. Li, H., Gao, K., Tu, H.: Variations in mode-specific valuations of travel time reliability and in-vehicle crowding: implications for demand estimation. Transp. Res. Part A Policy Pract. **103**, 250–263 (2017)
6. Pel, A.J., Bel, N.H., Pieters, M.: Including passengers' response to crowding in the dutch national train passenger assignment model. Transp. Res. Part A Policy Pract. **66**, 111–126 (2014)
7. Redman, L., et al.: Quality attributes of public transport that attract car users: a research review. Transp. Policy **25**, 119–127 (2013)
8. Qu, X., Meng, Q.: The economic importance of the straits of malacca and Singapore: an extreme-scenario analysis. Transp. Res. Part E Logistics Transp. Rev. **48**(1), 258–265 (2012)
9. Litman, T.: Valuing transit service quality improvement. J. Public Transp. **11**, 43–63 (2008)
10. Tirachini, A., Hensher, D.A., Rose, J.M.: Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. Transp. Res. Part A Policy Pract. **53**, 36–52 (2013)
11. Batarce, M., et al.: Valuing crowding in public transport systems using mixed stated/revealed preferences data: the case of Santiago. In: TRB 94th Annual Meeting Compendium of Papers, Washington DC (2015)
12. Gao, K., Shao, M., Sun, L.: Roles of psychological resistance to change factors and heterogeneity in car stickiness and transit loyalty in mode shift behavior: a hybrid choice approach. Sustainability **11**(17), 4813 (2019)

13. Wang, S., Zhang, W., Qu, X.: Trial-and-error train fare design scheme for addressing boarding/alighting congestion at CBD stations. Transp. Res. Part B Methodological **118**, 318–335 (2018)
14. Wang, S., Qu, X.: Station choice for Australian commuter rail lines: equilibrium and optimal fare design. Eur. J. Oper. Res. **258**(1), 144–154 (2017)
15. Bie, Y., et al.: Dynamic headway control for high-frequency bus line based on speed guidance and intersection signal adjustment. Comput. Aided Civil Infrastruct. Eng. **35**(1), 4–25 (2020)
16. Li, Z., Hensher, D.A.: Crowding and public transport: a review of willingness to pay evidence and its relevance in project appraisal. Transp. Policy **18**(6), 880–887 (2011)
17. Whelan, G.A. Crockett, J.: An investigation of the willingness to pay to reduce rail overcrowding. In: International Conference on Choice Modelling. Harrogate, England, Apr (2009)
18. Tirachini, A., et al.: Valuation of sitting and standing in metro trains using revealed preferences. Transp. Policy **47**, 94–104 (2016)
19. Vovsha, P., et al. Statistical Analysis of Transit User Preferences Including In-vehicle Crowding and Service Reliability. In: www.trb.org Compendium of papers TRB 93rd Annual Meeting. Washington, D.C (2014)
20. Shao, M., Li, T., Sun, L.: Survey method and model of passengers' cost perception of crowding level in bus. J. Tongji Univ. (Nat. Sci.) 07, 1031–1034 (2012)
21. Hirsch, L., Thompson, K.: I can sit but I'd rather stand: commuter's experience of crowdedness and fellow passenger behaviour in carriages on Australian metropolitan trains. (2011)
22. Wardman, M., Murphy, P.: Passengers' valuations of train seating layout, position and occupancy. Transp. Res. Part A Policy Pract. **74**, 222–238 (2015)
23. Gao, K., et al.: Heterogeneity in valuation of travel time reliability and in-vehicle crowding for mode choices in multimodal networks. J. Transp. Eng. Part A Syst. **144**(10), 04018061 (2018)
24. Li, Z., Hensher, D.A.: Crowding in public transport: a review of objective and subjective measures. J. Public Transp. **16**(2), 6 (2013)
25. Kuang, Y., Qu, X., Wang, S.: A tree-structured crash surrogate measure for freeways. Accid. Anal. Prev. **77**, 137–148 (2015)
26. Qu, X., et al.: Potential crash risks of expressway on-ramps and off-ramps: a case study in Beijing. China. Safety science **70**, 58–62 (2014)

Accessing the Influences of Weather and Environment Factors on Traffic Speed of Freeway

Danni Cao, Jianjun Wu, and Ziling Zeng

Abstract Traffic speed has been traditionally used as a measure of traffic performance. Predicting the traffic speed is fundamental for efficient traffic management and control strategy. This study explores the influences of freeway attributes, weather, and air condition on traffic speed. A quantitative model is also introduced to predict the traffic speed as per the identified influencing factors. Empirical data of traffic flow and potential influencing factors are collected from multiple sources for analysis and model calibration. The principal component analysis is firstly conducted to select the significant variables influencing the traffic speed. Afterward, a multiple linear regression model is calibrated to quantitatively model the impacts of different factors and investigate their weights. The results show that the attributes of freeway, the humidity of the area, the temperature, the horizontal visibility, the station maker, the air quality, and the PM quality have influences on the traffic speed. Among all of the variables, the weight of the existence of toll station is highest, indicating the largest influence on the traffic speed.

Keywords Traffic speed · Freeway · Variables selection · Model regression

1 Introduction

Traffic speed has been traditionally used as a measure of traffic performance. With the development of connected automated vehicles, intelligent transportation systems (ITS) are increasingly occupying an irreplaceable place [1]. Meanwhile, fast and accurate evaluation of traffic speed is a crucial requirement for the application of intelligent transportation systems [2, 3]. While the non-recurrent congestion

---

D. Cao · J. Wu (✉)
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, 100044 Beijing, China
e-mail: jjwu1@bjtu.edu.cn

D. Cao · Z. Zeng
Architecture and Civil Engineering, Chalmers University of Technology, 412 96 Goteborg, Sweden

induced by the incident hinder the stable and efficient of the traffic operation. Many researchers have paid much more attention on the traffic incidents, such as the traffic incident prevention and prediction, the propagation of congestion wave, and the effect of the traffic incident [4–8].

There is a strong positive correlation between the incident frequency and the traffic speed on the road [9]. Wang et al. found that an increase of 10 km/h on speed would make a 3% rise in incident frequency during peak periods on urban arterials of Shanghai [10]. Finch et al. also found that a one mile per hour increase in speed on rural roads would lead to a 4.9% increase in crash rate [11].

With the development of traffic monitoring systems, various types of data are available to adjust control measures in real time. It makes possible for the traffic management department to reduce congestion, avoid accidents, and improve traffic efficiency. Speed data has been typically collected by fixed-based sensors such as radar and loop detectors. When a vehicle passes the detector, the speed can be recorded. Based on the data of all the detectors on a freeway, the temporal and spatial variation of the speed of each location on the freeway can be explored.

The traffic condition can be influenced by many factors, such as the traffic flow, the freeway attributes, the weather, and the air condition. The speed may be completely different affected by various situations. While some factors may have internal correlations, and taking too many factors into consideration may increase the difficulty of data processing and biases in modeling the effect of different factors on speed. It is necessary to explore how different variables affect the speed. Principal component analysis can find a low-dimensional representation from a dataset with nice representation of variation in the dataset. Moreover, we apply multiple linear regression model to perform data regression, explore the relationship between various factors and speed, and quantitatively investigate the impacts of selected factors on traffic speed.

The contribution of this paper is twofold. Firstly, we introduce principal component analysis to reduce the complexity of dataset and select variables that may affect the traffic speed. Secondly, we understand the relationship between various factors and the traffic speed.

The remainder of this paper is organized as follows: Sect. 2 represents the detailed data description such as the speed record, weather data, freeway attributes, and air condition used in the case study. Section 3 introduces the principal component analysis and the multiple linear regression model. Section 4 conducts a numerical demonstration based on real freeway data in China. Section 5 concludes the study with limitations and future works.

2   Data Description

In this section, we introduce the detailed information of different kinds of data. We rely on the following four available data sources, which mainly are the freeway speed data, link attributes, the weather condition data, and the air condition data . The weather and air condition data are available from a meteorological data Web site (https://rp5.ru/).

Table 1 Description of weather and air condition variables

Variable	Description
Rain	It is rainy or not
Snow	It is snowy or not
T	The temperature at certain time
U	The humidity of the area
VV	The horizontal visibility of the area
AQI	The Air Quality Index of the area
PM2.5	The concentration of particulates smaller than or equal to 2.5 μm in diameter in the atmosphere

Speed data is measured by traffic detectors along with the freeway. And the free-way is divided into many links with different lengths. The speed data lasts from December 19, 2017, to December 23, 2017, and from January 2, 2018, to January 6, 2018. And the speed data is displayed in one minute.

Link attributes contain the information mainly about the length of the link, and whether there is a toll station or not.

Weather and air condition data information is displayed in 3 h and 1 h interval, respectively, listed in Table 1.

In our case study, we use speed data from December 19, 2017, to December 23, 2017, and from January 2, 2018, to January 6, 2018, and corresponding freeway link attributes, weather, and air condition. Totally, 23,070 speed records are obtained randomly, and corresponding link attributes, weather condition, and air condition are used. First, we focus on which is the real variables that influence the traffic speed. And then, we explore the relation between these selected variables and the traffic speed.

3   Methodology

In this section, we propose a systematic approach to explore how different vari-ables influence the traffic speed on freeway. Firstly, we introduce a method to reduce dimensionality, which is principal component analysis (PCA). We use PCA to select variables and reduce data complexity. Secondly, we apply the multiple linear regres-sion model to explore the relationship between selected variables and the traffic speed.

3.1 Principal Component Analysis

Speed is greatly influenced by many factors such as the weather, the air condition, and the road attributes. [12–18]. It does not mean that when we consider as many as factors, we will get a better result, as too many factors will increase the complexity of the problem. In many cases, there are correlations between different variables, indicating there is a certain overlap in the information of these variables on reflecting speed attributes. It is necessary to select representative factors influencing the speed. Principal component analysis can remove closely related variables from all the original variables. Data can be squeezed in this way that the variables can be reduced and replaced with principal dimensions without losing relevant information in the data.

Principal component analysis (PCA) aims to maximize the variance over a set of linear combinations. The basic mathematical formulation of PCA is given as follows:

(1) Calculate the average value.

Calculate the average value of different variables respectively, as shown in Eq. (1), and for all samples, the corresponding average value is subtracted as shown in Eq. (2).

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij} \tag{1}$$

$$x_{IJ} = x_{ij} - \bar{x}_j \tag{2}$$

(2) Normalization.

In some cases, the dimensions of the variables often differ greatly, so the impact of the dimensions should be eliminated before the principal component calculation which can be realized by Eqs. (3) and (4).

$$x_{ij}^* = \frac{x_{IJ}}{s_j} \tag{3}$$

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2} \tag{4}$$

(3) Calculate covariance matrix.

Take 3 variables as examples and the feature covariance matrix can be calculated by Eqs. (5) and (6).

$$C = \begin{bmatrix} \text{cov}(x_1, x_1) & \text{cov}(x_1, x_2) & \text{cov}(x_1, x_3) \\ \text{cov}(x_1, x_2) & \text{cov}(x_2, x_2) & \text{cov}(x_2, x_3) \\ \text{cov}(x_3, x_1) & \text{cov}(x_3, x_2) & \text{cov}(x_3, x_3) \end{bmatrix} \tag{5}$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) \tag{6}$$

(4) Calculate the eigenvectors and eigenvalues.

Find the eigenvalues $\lambda_1, \lambda_2, \lambda_3 \ldots$ of the covariance matrix and the corresponding eigenvectors, where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots$.

(5) Variables selection.

The basic idea of this step is to select the components that the eigenvalues are ranked first, and the cumulative contribution rate can reflect certain percentage information.

The principal component contribution rate can be obtained by $F(m)$ as shown in Eq. (7). And then, np principal components that explain certain percentage of the variance are selected.

$$F(m) = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{k=1}^{n_p} \lambda_i} \tag{7}$$

3.2 Correlation Analysis

Speed is impacted by various factors, such as the weather, the freeway link attributes, and the air condition. Terrible weather will have largely negative impact on speed. We want to explore how various variables impact the traffic speed and which is the most important factor. And we also want to study what form of speed will show the best fitted result with different factors.

We introduce the multiple linear regression model shown in Eqs. (8) and (9) for fitting data, where $\partial_0, \partial_1, \partial_2, \partial_3, \ldots \partial_n$ are the coefficients and $x_1, x_2, x_3, \ldots x_n$ are the variables selected by PCA which may affect the speed. $y_1, y_2, y_3, \ldots y_n$ are the speed of different records.

$$y = X\partial \tag{8}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ 1 & x_{21} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nn} \end{bmatrix}, \partial = \begin{bmatrix} \partial_0 \\ \partial_1 \\ \vdots \\ \partial_n \end{bmatrix} \tag{9}$$

4   Result and Analysis

In this section, we firstly use the real speed data of China freeway and select the real variables shown in Table 2 which influence the freeway speed with the method of PCA. The computer programs are written in Python 3.6. Secondly, we use a software called Stata to explore how these selected variables influence the traffic speed.

4.1   Variables Selection

According to former studies, we select some alternative variables listed in Table 2.

(1)   Calculate the average value
(2)   Normalization.

According to Table 3, the dimension of each variable is relatively different, and we need to normalize the raw data first.

(3)   Calculate covariance matrix

According to Eq. (5) and (6), the analysis result of covariance matrix is shown in Table 4.

(4)   Calculate the eigenvectors and eigenvalues

The result of eigenvalues is shown in Table 5.

(5)   Variables selection

Firstly, according to the eigenvectors and eigenvalues, we sort the eigenvalues shown in Table 6.

Secondly, we calculate cumulative principal component contribution rate $F(m)$ shown in Table 7.

Table 2 Description of different variables

Type	Variable
Link attribute	Link length (length)
Link attribute	Toll station marker (Ts)
Weather	Rain (rain)
Weather	Snow (snow)
Weather	Temperature (T)
Weather	U (U)
Weather	VV (VV)
Air	AQI (Aqi)
Air	PM2.5 (Pm)

Table 3 Mean value of different variables

Variable	Aqi	Pm	Rain	Snow	T	U	VV	Length	Ts
Mean	72.45	49.07	0.14	0.0035	3.81	70.48	11.60	851.51	1.02

Table 4 Covariance matrix

Variables	Aqi	Pm	Rain	Snow	T	U	VV	Length	Ts
Aqi	1	0.99	−0.27	−0.04	0.2	0.38	−0.56	0.23	0.02
Pm	0.99	1	−0.27	−0.03	0.21	0.38	−0.57	0.23	0.02
Rain	−0.27	−0.27	1	−0.02	0.16	0.38	−0.14	0.02	0.01
Snow	−0.04	−0.03	−0.02	1	−0.03	0.07	−0.06	0.04	0.01
T	0.20	0.21	0.16	−0.03	1	0.44	−0.22	0.07	0.02
U	0.38	0.38	0.38	0.07	0.44	1	−0.77	0.11	0.03
VV	−0.56	−0.57	−0.14	−0.06	−0.22	−0.77	1	−0.14	−0.02
Length	0.23	0.23	0.02	0.04	0.07	0.11	−0.14	1	−0.01
Ts	0.02	0.02	0.01	0.01	0.02	0.03	−0.02	−0.01	1

Table 5 Eigenvalues

Variable	Aqi	Pm	Rain	Snow	T	U	VV	Length	Ts
Eigenvalues	3.04	1.62	0.01	0.2	0.42	0.8	0.93	1.03	1

Table 6 Eigenvalues rank

Variable	Aqi	Pm	Rain	Snow	T	U	VV	Length	Ts
Rank	1	2	9	8	7	6	5	3	4

Table 7 Cumulative principal component contribution rate

Component	1	2	3	4	5	6	7	8	9
$F(m)$	0.34	0.52	0.63	0.74	0.85	0.94	098	0.99	1

Thirdly, based on the rank and cumulative contribution rate, the main variables are selected. We choose variables based on the selected variables which can indicate more than 98% information. So, we delete the last ranked variables, which are rain and snow.

4.2 The Influences of Weather and the Environment on Speed

Traffic waves show a clear regularity throughout the day, and the speed tendency is different in each time period. We select three representative time periods in hour unit which are 8:00, 12:00, and 16:00. We explore how different variables influence the traffic speed in different time intervals. Figure 1 shows the average speed in 24 h of random 10 different freeways. Different colored curves represent all-day traffic speed on different freeways. Take the orange one as illustration, it shows that the vehicle speed drops to the minimum value at 9:00 and then rises until 13:00; after that, the vehicle speed drops to approximately 10 km/h at 18:00 and finally rises until 24:00. It is obvious that different freeway links have a similar speed change tendency, and it has two lowest points in a day, which is consistent with the traffic flow characteristics.

Take 8:00 as an example, we find that the third power of speed is linear with different variables, which gives the best results; especially, U is logarithmic after repeated tests, as shown in Eq. (10). It means that a toll station on the freeway matters most on the traffic speed at 8:00 a.m. The horizontal visibility, the air humidity, and Aqi have negative effect on vehicle speed.

We have conducted many experiments and compare the statistical result. We listed several cases to decide the form of dependent variable and the U. The corresponding
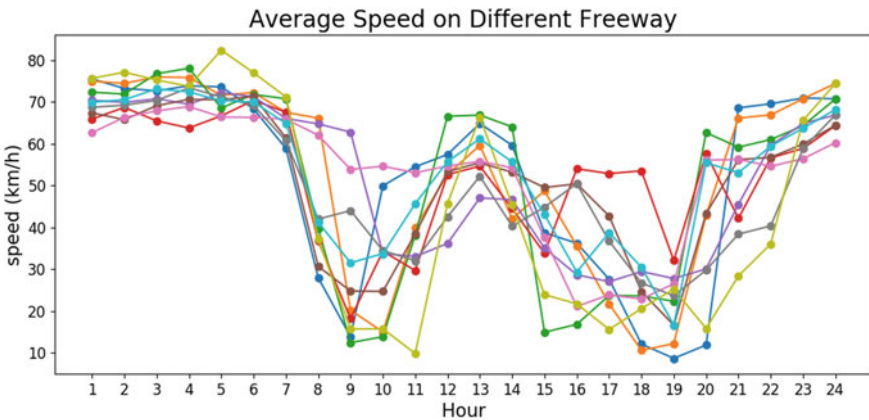
Fig. 1 Average speed on different freeways

Table 8 R-squared in different cases

Case	1	2	3	4	5	6	7	8
R-squared	0.243	0.243	0.29	0.29	0.3	0.3	0.29	0.29

Table 9 *P*-value of different variables in different cases

Case	1	2	3	4	5	6	7	8
Length	0	0	0	0	0	0	0	0
lgU/U	0.021	0.365	0.356	0.067	0.001	0	0	0
T	0	0	0	0	0	0	0	0
VV	0.018	0.001	0	0	0	0	0	0
Ts	0	0	0	0	0	0	0	0
Aqi	0	0	0	0	0	0	0	0
Pm25	0	0	0	0	0	0	0	0
cons	0	0	0	0	0	0	0	0

R-squared of different cases and the *P*-value of different variables are shown in Tables 8 and 9, respectively.

Case 1: v and U
Case 2: v and lg(U)
Case 3: v2 and U
Case 4: v2 and lg(U)
Case 5: v3 and U
Case 6: v3 and lg(U)
Case 7: v4 and U
Case 8: v4 and lg(U).

From Tables 8 and 9, we find that Case 6 gives the best fitted result and at the same time satisfies the 0.1% significance level.

$$v^3 = 150.625\text{Length} - 33{,}626.44\text{lgU} + 8913.506T$$
$$- 2073.108\text{VV} - 101{,}479.2\text{Ts} - 1809.955\text{Aqi}$$
$$+ 4872.766\text{Pm} + 551{,}091.6 \tag{10}$$

And we use the same method to explore the relationship between different variables and the traffic speed in other time intervals. Equations (11) and (12) are the results of 12:00 and 16:00. 20,133 data records and 23,069 data records are used, respectively. We find that in different time intervals, various variables show different impacts on speed, and in all of the fitted results, the Ts is the main factor. Whether there is a toll station on the freeway matters most on traffic speed.

$$v^3 = 152.812\text{Length} - 1000.196\text{lgU} + 4881.926T - 4112.655\text{VV}$$

$$- 113,300.2\text{Ts} + 2822.498\text{Aqi} - 1563.203\text{Pm} + 445,693.4 \qquad (11)$$

$$v^3 = 159.086\text{Length} + 610.529\text{lgU} + 6401.029T - 1961.01\text{VV}$$
$$- 106,682.9\text{Ts} + 2417.022\text{Aqi} - 1040.117\text{Pm} + 307,050.8 \qquad (12)$$

We find that the length, the temperature, and the Aqi have positive influences on speed. While it also indicates in different fitted equations, each coefficient fluctuates less except the U, and we guess it may be the problems with data quality, and it needs to be continued studying.

The research may be helpful for predicting traffic speed, and it also can be used for traffic management department to develop control measures such as variable speed limits in different scenarios.

5 Conclusion

This study explores the influences of freeway attributes, weather, and air condition on traffic speed. A quantitative model is also introduced to predict the traffic speed as per the identified influencing factors. Empirical data of traffic flow and potential influencing factors are collected from multiple sources for analysis and model calibration. In order to reduce the complexity of data processing, we firstly introduce principal component analysis. The PCA can reduce data dimension and select irrelevant variables which may have impact on traffic speed. Secondly, we apply multiple linear regression model to understand how different factors influence traffic speed and explore the form of traffic speed that fits better with different variables. Finally, empirical data of China freeway is used to illustrate the feasibility of the method. For the traffic speed of the freeway link has strong cycle fluctuations throughout the day, we fit the data in hour. We find that the length of freeway, the humidity of the area, the temperature, the horizontal visibility, the station maker, the air quality, and the pm quality have influences on vehicle. Among all of the variables, the weight of the existence of toll station is highest, indicating the largest influence on the traffic speed.

In spite of the promise of this novel approach for measuring how different variables affect traffic speed, we must note some shortcomings which should be addressed in the future work. Firstly, in our research, we only select some variables which may ignore other factors' impact. In the future research, we will take more variables into consideration. Secondly, we will introduce other models for data regression and compare the results of different models. Thirdly, we will make real-time predictions of speed based on this research.

References

1. Li, X., Ghiasi, A., Xu, Z., Qu, X.: A piecewise trajectory optimization model for connected automated vehicles: exact optimization algorithm and queue propagation analysis. Transp. Res. Part B Methodol. **118**, 429–456 (2018)
2. Qu, X., Zhou, M., Yu, Y., Lin, C.T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach. Appl. Energ. **257**, 114030 (2019)
3. Zhou, M., Yu, Y., Qu, X.: Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: a reinforcement learning approach. IEEE Trans. Intell. Transp. Syst. **21**, 433–443 (2019)
4. Zhou, M., Qu, X., Li, X.: A recurrent neural network based microscopic car following model to predict traffic oscillation. Transp. Res. Part C. **84**, 245–264 (2017)
5. Kuang, Y., Qu, X., Wang, S.: A tree-structured crash surrogate measure for freeways. Accid. Anal. Prev. **77**, 137–148 (2015)
6. Qu, X., Meng, Q., Li, S.: Ship collision risk assessment for the Singapore Strait. Accid. Anal. Prev. **43**, 2030–2036 (2011)
7. Xu, C., Yang, Y., Jin, S., Qu, Z., Hou, L.: Potential risk and its influencing factors for separated bicycle paths. Accid. Anal. Prev. **87**, 59–67 (2016)
8. Gao, K., Tu, H., Sun, L., Sze, N. N., Song, Z., Shi, H.: Impacts of reduced visibility under hazy weather condition on collision risk and car-following behavior: Implications for traffic control and management. Int. J. Sustain. Transp. 1–8 (2019)
9. Oh, C., Oh, J., Ritchie, S., Chang, M.: Real-time Estimation of Freeway Accident Likelihood, 80th Annual Meeting of the Transportation Research Board. Washington, D.C, Washington, DC (2001)
10. Wang, X., Song, Y., Yu, R., Schultz, G.G.: Safety modeling of suburban arterials in Shanghai, China. Accid. Anal. Prev. **70**, 215–224 (2014)
11. Finch, D.J., Kompfner, P., Lockwood, C.R., Maycock, G.: Speed, speed limits and crashes. Project record S211G/RB/project report PR 58. Transport research laboratory TRL, Crowthorne, Berkshire. (1994)
12. Ahmed, M.M., Abdel-Aty, M.A.: The viability of using automatic vehicle identification data for real-time crash prediction. IEEE Trans. Intell. Transp. Syst. **13**, 459–468 (2012)
13. Joo, S., Oh, C., Lee, S., Lee, G.: Assessing the impact of traffic crashes on near freeway air quality. Transp. Res. Part D Transp. Environ. **57**, 64–73 (2017)
14. Rahman, A., Lownes, N.E.: Analysis of rainfall impacts on platooned vehicle spacing and speed. Transp. Res. Part F Traffic Psychol. Behav. **15**, 395–403 (2012)
15. Wang, Y., Luo, J.: Study of rainfall impacts on freeway traffic flow characteristics. Transp. Res. Procedia **25**, 1533–1543 (2017)
16. Qu, X., Zhang, J., Wang, S.: On the stochastic fundamental diagram for freeway traffic: model development, analytical properties, validation, and extensive applications. Transp. Res. Part B **104**, 256–271 (2017)
17. Qu, X., Wang, S., Zhang, J.: On the fundamental diagram for freeway traffic: a novel calibration approach for single-regime models. Transp. Res. Part B **73**, 91–102 (2015)
18. Xu, Z., Wei, T., Easa, S., Zhao, X., Qu, X.: Modeling relationship between truck fuel consumption and driving behavior using data from internet of vehicles. Compu. Aided Civil Infrastruct. Eng. **33**(3), 209–219 (2018)

Multivariate Time Series Analysis Using Recurrent Neural Network to Predict Bike-Sharing Demand

Kanokporn Boonjubut and Hiroshi Hasegawa

Abstract The bike-sharing service system is a service that allows a customer to rent a bike from a bike-sharing station and then return it to another bike-sharing station in a short time after they reach their destination. Thus, the impact of the bike distribution system based on the frequency of bike usage needs to be assessed. The bike-sharing system operator needs to predict the demand to accurately know how many bikes are needed in every station so as to assist the planner in the management process of the bike-sharing stations. This paper proposes an efficient and accurate model for predicting the bike-sharing service usage using various features of a machine learning algorithm. We compared the exiting techniques for the sequential data predicting of artificial intelligence for time series data and analysis. Then, we considered the use of the multivariate model with a recurrent neural network (RNN), a long short-term memory (LSTM), and a gated recurrent unit (GRU). In addition, we considered combining the LSTM and GRU methods together to improve the model's effectiveness and accuracy. The results showed that all the RNNs, including the LSTM, GRU, and the model combining the LSTM and GRU, are able to achieve high performance using the mean square mean absolute, mean squared error, and root mean square error. However, the mixed LSTM–GRU model accurately predicted the demand in this case.

Keywords Artificial intelligence · Time series analysis · Predict demand

1 Introduction

Recently, bike-sharing systems (BSSs) have become very popular, as the growing numbers of people and vehicles have increased traffic congestion and led to many

K. Boonjubut (✉)
Functional Control Systems, Shibaura Institute of Technology, Saitama, Japan
e-mail: nb19103@shibaura-it.ac.jp

H. Hasegawa
Machinery and Control Systems, Shibaura Institute of Technology, Saitama, Japan

environmental concerns. Also, bike sharing has been a growing demand for transportation services [1]. Moreover, BSSs are sustainable transportation alternatives to private transport, as they do not lead to carbon emissions, traffic congestion, or the use of non-renewable resources [2]. However, the service quality has a great impact on customer satisfaction, which affects the increase in the number of customers, the service's popularity, and the overall economic performance of the bike-sharing companies.

Bike-sharing services have also had much attention in metropolitan areas and tourist areas. In general, a BSS allows registered customers to request a ride after indicating the pickup and drop-off times and locations. Afterward, a customer will be able to find an available bike and then park it at any station after reaching their destination. Thus, BSSs need to maintain the optimal number of bikes and parking spots at each station. Predictive analytics is a collective term that refers to the methods of business analytics and data science, and it is used in the prediction of the necessary variables that are to be studied as continuous variables, such as predicting sales in a new business. Whether starting a new business or adding a new cycle to the goals of a business operation, predictive analytics can lead to maximum profits and minimal costs. Several studies have predicted the usage of bikes in future transportation systems. Singhvi et al. [3] have presented a bike demand model using regression models to predict bike trips during morning rush hours on weekdays. Zeng et al. [4] have proposed a mobility and prediction model for BSSs, and they used statistical methods to predict the spatiotemporal shift of bikes between stations.

In most cases, the use of bike sharing is often represented by a time series data that need to be predicted. At present, there is no information about using machine learning for the multivariate time series analysis that can be used to predict the variables of BSSs. Therefore, this paper proposes a model that can predict the number of bikes to be used at each bike-sharing station. The model mainly predicts the demand for bikes using machine learning techniques such as recurrent neural networks (RNNs) [5]. Ashqar et al. [6] have proposed a bike-sharing station model using a multivariate regression algorithm, as the partial least squares regression random was applied in their study, including RNN, long short-term memory (LSTM), and gated recurrent unit (GRU). In this paper, we used a case study from an open source in New York City to improve the system's management and create an accurate model for predicting the use of bike sharing, which will be able to assist planners.

2 Methodology

2.1 Data Preprocessing

We used data from a BSS in New York City (Citi Bike). We downloaded the bike-sharing usage data from January 1, 2011, to December 31, 2012, from Citi Bike's official Web site and acquired the hourly weather data from the UCI bike-sharing

dataset [7]. In this model, we only used the numbers of rents and returns, information on the stations, and numbers of passengers, and the dataset and variables are shown in Table 1. According to the raw data, we divided by each day and collected the dataset, which was ID instead of the station name and the number of users per hour added together. We also considered other various factors, such as historical weather data and days of the week. Afterward, we converted the data into a training dataset and tested it with a ratio of 70:30, respectively, where one sequence contained 30 consecutive daily data inputs. Then, the data were transformed using MinMax scaler [8] by scaling each feature to a given range. This scaler works in scales across input variables that may have different units, which can be used to create the required model for predicting the future demand for bike-sharing usages.

From the dataset, we created a coefficient correlation diagram using scikit-learn and Keras as shown in Fig. 1, and the results showed the correlation of each variant

Table 1 Dataset and used variables

Variable	Describe
Date time	Indicates date time including the date, month, year, and time information
Weekday	Indicates variables it is a weekday or weekend
Holiday	Indicates variable that government holiday is not it
Season	Divided into four seasons: spring, summer, fall, and winter
Weather	Divided into four types: transparent, mist and cloudy, rain, and snow
Temperature	Temperature in Celsius
Humidity	The values of humidity
Wind speed	The values of wind speed
Count	The number of bikes used

Fig. 1 Correlation coefficient diagram

(feature), where each feature depends on the temperature, season, humidity, wind speed, weekday, holiday, and count.

2.2 Models

In this section, we present a basic introduction of the techniques for the sequential data forecasting/predicting for time series data, where the analysis is one in statistical. We also propose a model to predict the future usages of bike-sharing services. Many articles were presented with univariate datasets. However, the data of real systems do not only have one or two different variables (bivariable), but they have three or more variables, which is why multivariable analysis is needed, for which a multivariable dataset is required.

The data in a series of time periods that involve the trend, seasonality, and the cycle are called the time series data, and this method can be applied to accurately solve this case study.

According to various predicting/forecasting methods performed in the application, the advantages of artificial intelligence, including machine learning techniques, can decrease the errors and increase the chances of attaining higher accuracy. Generally, an RNN, an LSTM, a GRU, and a combination of each model can be used. However, a new architecture model combing LSTM and GRU is used in this paper.

2.2.1 Recurrent Neural Network

A model in an artificial neural network algorithm that can recognize patterns such as time series data, natural language processing, and video recognition in addition to other patterns is called the RNN [9], and it is the process of the sequential data, where the previous hidden state is used to calculate the current hidden state and the current hidden state is used to estimate the next period's state.

$$h_t = (x_t W + h_{t-1} U) \tag{1}$$

$$o_t = (h_t V), \tag{2}$$

where t is the current step, h is the hidden state, x is the input, o is the output, σ is the activation function, and W, U, V are the weight matrices connecting x_t (input) to the previous hidden state h_{t-1} and the current hidden state h_t.

Therefore, the RNN is suitable for the sequence data or the time series data, where the RNN can be used with data that are not different in distance in a sequence. The problem with the transition function is the vanishing gradient influent difficulty learn [10–12].

2.2.2 Long Short-Term Memory (LSTM)

A kind of an RNN is used to process the sequential data, where the LSTM is developed based on the RNN [13, 14] while using the cell state and the hidden stage to store and access data in the next stage so as to prevent the disappearance of the gradient problem. They are deciding for three gates to consist of the forgetting gate, which considers the unnecessary information, and the input gate is saved in the cell stage. The output gate is the data transmission into an output.

$$f_t = \left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{3}$$

$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{4}$$

$$\widetilde{C}_t = \tanh\left(W_c \times \left[h_{t-1}, x_t\right] + b_c\right) \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{6}$$

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{7}$$

$$h_t = o_t * \tanh(C_t), \tag{8}$$

where t is the current step, x is the input, o is the output, W is the weight matrix, and b is the bias. f_t is the forget gate at a time t, i_t is the input gate, C_t is the cell state of the internal memory at a time t, h_t is the hidden state, and $\widetilde{C}_t$ is the candidate hidden state that can be used in the next state, which decides to remember or forget the input data.

2.2.3 Gated Recurrent Unit (GRU)

GRU is a developed unit based on the LSTM in a recurrent network. The GRU process is a gate adjustment in the LSTM into the reset gate. Then, the gate is updated to consider how much enough data in the previous stage merges into the input and forget gate that can be saved more computation resources than LSTM.

The hidden state of the previous stage is connected to the reset gate, and the GRU exposes all the memory, as it does not have the output gate. It consists of two gates: the reset gate and the update gate.

$$z_t = \sigma\left(W_z\left[h_{t-1}, x_t\right]\right) \tag{9}$$

$$r_t = \sigma\left(W_r\left[h_{t-1}, x_t\right]\right) \tag{10}$$
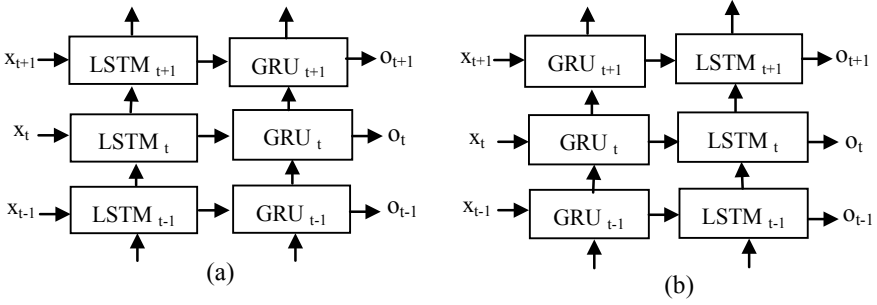
Fig. 2 Structures of the mixed LSTM–GRU

$$\tilde{h}_t = \tanh\big([r_t * h_{t-1}, x_t]\big) \tag{11}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t, \tag{12}$$

where r is the reset gate, z is the update gate that decides how much the unit updates its hidden state, h is the hidden state, $\tilde{h}$ is the candidate hidden state, σ is the activation function, W and U represent the weight matrix, t is the time period, and x is the input vector.

2.2.4 The Combination of LSTM and GRU

A model combining an LSTM and a GRU is proposed [15], which are assembled and connected in time, and a parameter propagation method using the LSTM or the GRU sequence is considered. The cell and hidden state are converted to the next unit, where the long-term memory affects the current epoch output. As for the multiple layers between the LSTM and the GRU, the first LSTM or GRU layer calculates the hidden units using the inputs. Then, the second LSTM or GRU layer calculates the output of the hidden units, but the different parameters are independently calculated. At last, the neural network calculates the loss function and tries to optimize it. Figure 2 shows the structure of the combined LSTM and GRU.

2.2.5 Experimental

In this paper, the dataset of the input matrix consists of temperature, season, humidity, wind speed, weekdays, holidays, and count to predict future bike-sharing usage. In the experiment, we used the hyperparameter as presented in Table 2. The performance of the results became stable after selecting each of the variables. The time step is 30 because the pattern of the data was acquired for every 30 days together. Furthermore, we used the same hyperparameter to build the proposed model. In the next step, each

Table 2 Description of the variables

Parameter	Value
Epoch	100
Batch size	100
Hidden node	100
Hidden layer	1–2
Time step	30

training step is processed. Then, the predicted results are estimated according to the bike-sharing usage of each model.

2.2.6 Evaluation

We used the mean square mean absolute (MAE), mean squared error (MSE), and root mean square error (RMSE) methods to measure the effectiveness and accuracy of different models and to evaluate the bike-sharing usage prediction using the proposed model with multivariate time series analysis. The predicting error was calculated as follows, where n is the testing sample number, y is the real data of the bike-sharing usage, and $\tilde{y}$ is the corresponding prediction of the bike-sharing usage.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \tag{13}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \tilde{y}_i| \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_i - y_i)^2} \tag{15}$$

3 Results

We used a different model to predict the demand and compare the results. We used the MAE, MSE, and RMSE methods to measure the performance of the proposed model. Table 3 shows the predicted errors of the predictions of the bike-sharing usage and raw bike-sharing usage. The experimental results showed that the mixed LSTM–GRU model had the least error values, as its MAE, MSE, and RMSE are 0.0420, 0.0040, and 0.0630, respectively.

Table 3 Predicted errors of the bike-sharing usage

Model	MAE	MSE	RMSE
RNN	0.0744	0.0055	0.0505
LSTM	0.0510	0.0510	0.0786
GRU	0.0494	0.0053	0.0735
LSTM–GRU	0.0420*	0.0040*	0.0630*
GRU–LSTM	0.0470	0.0050	0.0708

4 Conclusion and Future Work

In this paper, time series data and a machine learning technique were used to predict the demand for bike-sharing usage. Using historical data, including the season, weather, temperature, humidity, and wind speed with the time series analysis, we compared the performance of different prediction models that used an RNN, an LSTM, a GRU, a mixed LSTM–GRU, and a mixed GRU–LSTM. The results showed that the prediction of the demand for bike-sharing usage with the mixed LSTM–GRU model was the most accurate. Thus, the findings of this study can assist the planners of bike-sharing companies in the distribution of bikes in bike stations. For future work, we will consider the effects of each variable on the bike usage rate and will further develop the predicted results using other methods such as deep learning.

References

1. Shaheen, S.A., Guzman, S., Zhang, H.: Bike-sharing in Europe, the Americas, and Asia: past, present, and future. Transp. Res. Rec. **2143**(1), 159–167 (2010)
2. Ghosh, S., Varakantham, P., Adulyasak, Y., Jaillet, P.: Dynamic repositioning to reduce lost demand in bike sharing systems. J. Artif. Intell. Res. **58**, 387–430 (2017)
3. Singhvi, D., Singhvi, S., Frazier, P.I., Henderson, S.G., O'mahony, E., Shmoys, D.B., Woodard, D.B.: Predicting bike usage for New York city's bike sharing system. In: Dilkina, B., Ernon, S., Hutchinson, R.A., Sheldon, D. (eds.) AAAI Workshop: Computational Sustainability. AAAI Press (2015)
4. Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., Moscibroda, T.: Mobility modeling and prediction in bike-sharing systems. In: Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 165–178. Association for Computing Machinery, Singapore (2016)
5. Pana, Y., Zhenga, R.C., Zhanga, J., Yaob, X.: Predicting bike sharing demand using recurrent neural networks. Procedia Comput. Sci. **147**, 562–566 (2019)
6. Ashqar, H.I., Elhenawy, M., Almannaa, M.H., Ghamem, A., Rakha, H.A., House, L.: Modeling bike availability in a bike-sharing system using machine learning. In: 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp. 374–378. IEEE, Italy (2017)
7. Bike-sharing demand. http://www.capitalbikeshare.com/system-data. Last accessed 20 Nov 2019
8. Bishop, C.M.: Neural networks for pattern recognition, 1st edn. Oxford University Press, Oxford, USA (1996)

9. Zaremba, W., Sutskever, I., Vinyals, O.: Recurrent neural network regularization. CoRR arXiv:1409.2329 (2014)
10. Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J.: Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. In: Kolen, J.F., Kremer, S. (eds.) A Field Guide to Dynamical Recurrent Networks, pp. 237–243. IEEE Press (2001)
11. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
12. Cho, K., Merrienboer, B.V., Gulechre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734. Association for Computational Linguistics, Doha, Qatar (2014)
13. Sak, H., Senior, A.W., Beaufays, F.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association, pp. 338–342. ISCA, Singapore (2014)
14. Olah, C.: Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs. Last accessed 20 Nov 2019
15. Jiang, C., Chen, S., Bo, Y., Wang, Y., Sun, Z.: Performance improvement of GPS/SINS ultra-tightly integrated navigation system utilizing a robust cubature Kalman filter. J. Aeronaut. Astronaut. Aviat. **49**(1), 49–55 (2017)

Influencing Factor Analysis of Logistics Service Satisfaction in China: A Binary Logit Model Approach

Wen Xu, JiaJun Li, and Bin Shen

Abstract With the continuous growth of the number of Internet users and the continuous expansion of the online retail market in China, consumers have put forward higher requirements for logistics services. Thus, this study uses a binary logit model to study China's logistics services from the perspective of customers, and the main factors affecting customer satisfaction with logistics services were explored. First, questionnaire was scientifically designed, and a total of 356 samples was collected in the online survey, of which 310 samples considered to be valid, and the questionnaire effective rate was 87.1%. Second, based on the survey data, the reliability and validity of the questionnaire were tested and factor analysis was performed. The results showed that the reliability and validity of the questionnaire were good, and the factor structure of the questionnaire could meet the needs of this study. Finally, the binary logit model was used to analyze the main factors affecting customer satisfaction with logistics services. The results show that facilities, convenience, reliability, empathy, economics and timeliness have a significant impact on logistics service satisfaction, and among them, facilities, economics and convenience are the most important factors. The research results can effectively improve China's overall logistics service level and have strong practical significance.

Keywords Logistics services · Questionnaire · Factor analysis · Binary logit model

W. Xu · J. Li (✉)
School of Management, Northwestern Polytechnical University, Xi'an 710072, China
e-mail: 18078115229@126.com

W. Xu
Guangxi Computing Center CO., Ltd., Guangxi Communications Investment
Group CO., Ltd., Nanning 530022, China

B. Shen
School of Civil Engineering and Transportation, South China University of Technology,
Guangzhou 510641, China

1   Introduction

In recent years, the expansion of online retail sales has increased rapidly due to the development of information technology in China [1, 2]. According to the *44th Statistical Report on Internet Development in China* issued by China Internet Network Information Center (CNNIC), up to June 2019, the number of online shopping users in China has reached 639 million, an increase of 28.71 million compared with the end of 2018 [3]. Fast-growing online retail market provides customers more convenience but places higher demands on logistics services as well. Logistics services usually refer to logistics management from one point to another to meet the requirements of customers or companies [4], which is an important component in online retail market and significantly impacts online retailers' revenue and profitability [5].

Many scholars have carried out a lot of research on logistics services. Sudrajat constructed a conceptual model, for improving the strategic competitiveness of Indonesian logistics service companies [6]. Wang investigated consumers' behavioral responses to innovative the last-mile delivery services in the emerging e-commerce market of Singapore [7]. Thao explored different stakeholder perceptions of logistics service quality in Hai Phong, Vietnam, and provides managers with insights on what logistics service quality factors to address to enhance customers' perceptions [8]. Lin used factor analysis method to identify crucial business and logistics skill factors in Taiwan and analyzed the different perception between international logistics service providers and logistics academics [9]. Limbourg examined the logistics service quality in Da Nang City by scales developed according to SERVQUAL instrument [10].

Customer satisfaction is very important for online retail market and logistics companies. Studies have shown that customer satisfaction with logistics services has an important impact on their online shopping behavior [11, 12]. In order to evaluate customer satisfaction, large numbers of research have been carried out. Briggs examined the influence of positional performance and velocity performance on the third-party logistics service satisfaction based on an online survey [13]. Rahmat studied the influence of cultural background between individualist and collectivist countries on customer logistics service satisfaction [14]. Stopka put forward some proposals of methodology for measuring the customer satisfaction with the logistics services based on the directly and indirectly measurable characteristics [15]. Tontini study the nonlinear impact of quality dimensions of third-party logistics (3PL) services on customer satisfaction and loyalty, by using penalty and reward contrast analysis method [16]. Duc studied the service quality–customer satisfaction link in the port logistics service industry of a developing and transitional economy by structural equation modeling technique [17].

The 44th Statistical Report on Internet Development in China pointed that the number of Internet users in China has reached 854 million, and the Internet penetration rate has reached 61.2%, up to June 2019. In the context of the growing number of Internet users and the expanding of online retail market, it is necessary to re-investigate customer logistics service satisfaction in the Chinese market. This paper

uses the method of factor analysis and binary logit model to study the satisfaction of logistics services in China and can effectively make up for the research gap, which is significant for improving the competitiveness of logistics companies and logistics services in China.

2 Method

2.1 Questionnaire Design

Based on the previous studies of customer satisfaction with logistics, this research developed a questionnaire with 31 items to study the logistics satisfaction in China under the context of the rapid development of online retail market. The questionnaire was made of two parts: The first part was an investigation of the demographic information included: age, gender, income and the frequency of logistics services used per week; the second part was an investigation on the logistics satisfaction which had 21 items, which included seven topic areas: reliability, timeliness, economics, convenience, empathy, facilities and satisfaction (see in Table 1). The five scale of Likert was applied in the logistics satisfaction questionnaire; each measurement variable varies from 1 to 5, among which 1 means strongly agree, 2 means agree, 3 means uncertain, 4 means disagree and 5 means strongly disagree.

2.2 Survey Implementation

This survey was conducted from 2019-12-06 to 2019-12-20 by means of an online survey. The online questionnaire was published on Wenjuanxing (https://www.wjx.cn/), one of the most popular questionnaire survey Web sites in China. This survey was completely anonymous for the sake of privacy issues. Respondents were encouraged to participate in the survey by giving them five Chinese yuan after the survey; the average survey completion time was 3 min. Altogether, 356 people took part in the study. After deleting unreliable cases by controlling for missing values, answering time for the entire survey (answering time less than 90 s), 310 people formed the final sample, resulting in a sample validity rate of 87.1%.

2.3 Survey Sample Structure

In terms of the demographic information of the participants, as can be seen in Table 2, 157 samples (50.6%) are male, while 153 (49.4%) are female; the ratio between male and female approaches 1:1. More than half (163, 52.6%) of the participant's age is

Table 1 Logistics satisfaction questionnaire

Variable	Codes	Items
Reliability	Rel1	Your package can be delivered accurately and will not be falsely taken
	Rel2	Your package was intact when arrives in your hands
	Rel3	You have not lost you package before it delivered to your hand
	Rel4	Your package was safe and reliable during the delivery process
Timeliness	Tim1	Your package will be delivered to your hand on time as expected
	Tim2	On special holidays (such as Double Eleven), your courier can deliver in time
	Tim4	When you return or exchange, the courier can pick up the goods in time
Economics	Eco1	You are very satisfied with the cost of current logistics services
	Eco2	You think that the pricing of logistics services is very reasonable
	Eco3	You think the current cost-effectiveness of logistics services is relatively high
Convenience	Con2	Express distribution of pickup points is reasonable, and it is convenient
	Con3	The daily pickup time is long enough and flexible for you
	Con4	There are many pickup methods available for you to choose
Empathy	Emp1	The courier's attitude is very friendly when picking up
	Emp2	The logistics company can provide services corresponding your needs
	Emp3	The logistics company can provide me with a lot of personalized services
	Emp4	The logistics company can meet my needs very well in the service process
Facilities	Fac1	The facilities and equipment of logistics companies are modernization now
	Fac2	The logistics company staffs can provide logistics standardization services
	Fac3	The appearance and clothing of logistics company employees are very good
Satisfaction	Sat	You are very satisfied with the service provided by the logistics company

between 26 and 40, and people in this age group are also major users of logistics services. A total of 141 (45.5%) participants' income is above 6000 yuan per month. Almost half of the participants (151, 48.7%) use logistics services 2–3 times a week.

Table 2 Descriptive statistics of the demographic information of the participants

Demographic variables	Value	Frequency	Percentage	Cumulative percentage
Gender	Male	157	50.6	50.6
	Female	153	49.4	100
Age	<18	2	0.6	0.6
	18–25	31	10.0	10.6
	26–40	163	52.6	63.2
	41–55	96	31.0	94.2
	>55	18	5.8	100.0
Income	<1500	15	4.8	4.8
	1500–2999	49	15.8	20.6
	3000–4999	54	17.4	38.1
	4500–5999	51	16.5	54.5
	>6000	141	45.5	100.0
Frequency of logistics services used per week	≤ 1	109	35.2	35.2
	2–3	151	48.7	83.9
	4–6	25	8.1	91.9
	>7	25	8.1	100.0

3 Result

3.1 Reliability Test of the Logistics Satisfaction Questionnaire

Reliability refers to the consistency and stability of the data measured by the questionnaire. The most widely used reliability index is Cronbach's α coefficient, and its value in scale should better be above 0.7 [18]. Used SPSS 22.0 software to analyze the reliability of the logistics satisfaction questionnaire, results are shown in Table 3. From Table 3, it could be observed that the value of each factor was greater than 0.7.

Table 3 Cronbach's α coefficient of the logistics satisfaction questionnaire

Variable	Quantity	Cronbach's α coefficient
Reliability	4	0.780
Timeliness	3	0.761
Economics	3	0.833
Convenience	3	0.840
Empathy	4	0.888
Facilities	3	0.853
Total	20	0.934

At the same time, Cronbach's α coefficient in the total scale is 0.934. Those values of Cronbach's α coefficient indicated that the logistics satisfaction questionnaire was reliable.

3.2 Factor Analysis of the Logistics Satisfaction Questionnaire

Factor analysis method was adopted to analyze the structure of the logistics satisfaction questionnaire. First, the KMO value was analyzed and the Bartlett's test of sphericity was carried out by using SPSS 22.0. The value of KMO was 0.922, and the result of Bartlett's test of sphericity was significant (<0.01), which means we can apply explanatory factor analysis to explore the structure of the logistics service questionnaire. The results of the analysis are presented in Table 4; there were six common factors with the variance explanation rate valuing 73.766%. All the factor loads are greater than 0.5 (0.520–0.868), which indicated that the logistics satisfaction questionnaire had good validity [19].

3.3 Construction of the Logistics Satisfaction Logit Model

This study utilized a binary logit model approach to estimate the main factors affecting logistics service satisfaction. First, the binary logit model requires the dependent variable to be a binary variable so that the logistics service satisfaction variable was grouped into two categories [20, 21], that is "very satisfied" and "satisfied" that were grouped into "satisfied" with a value of 1, and the results of "unclear," "unsatisfied" and "very dissatisfied" were grouped into "unsatisfied" with a value of 0. Then, use the binary logistic regression models to estimate the relationship between the logistics satisfaction (satisfied/unsatisfied) and the explanatory variables of facilities, convenience, reliability, empathy, economics and timeliness. The logistic regression equation can be expressed as Formula (1):

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \qquad (1)$$

In this formula, P represents the probability of the logistics service satisfaction that is "satisfied," $1 - P$ represents the probability of the logistics service satisfaction that is "unsatisfied," β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients of the explanatory variables and $X_1, X_2, \ldots, X_n$ are the explanatory variables.

To assess the goodness of fit of the fitted model, the Hosmer and Lemeshow test was performed by using SPSS 22.0, and the result is shown in Table 5. The result shows that chi-square $= 13.317$ and P-value >0.101 indicated that there is

Table 4 Factor analysis results of the logistics satisfaction questionnaire

Items	Mean	Factors loading	Variance explained (%)	Cumulative variance explained (%)
Factor 1: facilities			14.580	14.580
Fac1:	3.76	0.827		
Fac2:	3.72	0.761		
Fac3:	3.71	0.716		
Factor 2: convenience			13.223	27.804
Con3:	3.85	0.794		
Con2:	3.90	0.773		
Con1:	3.82	0.698		
Factor 3: reliability			12.689	40.493
Rel3:	4.25	0.805		
Rel1:	4.29	0.768		
Rel2:	4.01	0.637		
Rel4	3.99	0.616		
Factor 4: empathy			12.056	52.549
Emp4:	3.80	0.615		
Emp2:	3.89	0.614		
Emp1:	4.00	0.610		
Emp3	3.67	0.607		
Factor 5: economics			12.015	64.564
Eco2:	3.49	0.868		
Eco1:	3.64	0.806		
Eco3:	3.65	0.684		
Factor 6: timeliness			9.202	73.766
Tim4:	3.86	0.800		
Tim1:	3.82	0.618		
Tim2:	3.56	0.520		

Table 5 Goodness of fit of the binary logistic regression model

Step	Chi-square	df	P-value
1	13.024	8	0.111

no significant difference between the actual value and the predicted value of the dependent variable, and the model was fitted well [22].

Table 6 summarizes the estimated result of the fitted binary logistic regression model, and according to the fitted result, the final form of the binary logistic regression model can be described as Formula (2):

Table 6 Estimated result of the fitted binary logistic regression model

	Coefficient	S.E.	Wald	df	P-value	OR	95% C.I. (OR)	
							Lower	Upper
Facilities	1.522	0.230	43.820	1	0.000	4.583	2.920	7.192
Convenience	0.939	0.183	26.474	1	0.000	2.558	1.789	3.659
Reliability	0.781	0.188	17.332	1	0.000	2.185	1.512	3.156
Empathy	0.721	0.168	18.303	1	0.000	2.056	1.478	2.860
Economics	1.003	0.196	26.163	1	0.000	2.727	1.857	4.006
Timeliness	0.652	0.174	14.105	1	0.000	1.919	1.366	2.697
Constant	1.967	0.237	69.078	1	0.000	7.150		

$$\text{logit}(P) = 1.967 + 1.522X_1 + 0.939X_2 + 0.781X_3$$
$$+ 0.721X_4 + 1.033X_5 + 0.652X_6 \tag{2}$$

In this formula, X_1 represents the value of facilities, X_2 represents the value of convenience, X_3 represents the value of reliability, X_4 represents the value of empathy, X_5 represents the value of economics and X_6 represents the value of timeliness.

From the estimated model, it was found that facilities, convenience, reliability, empathy, economics and timeliness have a significant impact on logistics service satisfaction, and among them, facilities, economics and convenience are the most important factors. Customers who are satisfied with the logistics facilities are about 358% more likely to rate the logistics service as "satisfied" compared to customers who are dissatisfied with the facilities (OR = 4.583; $p = 0.000$); this means that when logistics companies adopt more advanced logistics facilities, customers tend to be satisfied with the logistics service. Similarly, customers who are satisfied with the economics of the logistics service are about 273% more likely to rate the logistics service as "satisfied" compared to customers who are dissatisfied (OR = 2.727; $p = 0.000$) and indicated that customer satisfaction with logistics services marginally depends on how much they spend on logistics services. With respect to the convenience of the logistics services, customers who are satisfied with the level of convenience of the logistics services are about 256% more likely to rate the logistics service as "satisfied" compared to customers who are dissatisfied (OR = 2.558; $p = 0.000$) and reflected that convenience is an important factor influencing customer satisfaction with logistics services.

4   Conclusion

This study analyzed the main factors affecting customer logistics service satisfaction in the context of the growing online retail market in China, by using the binary logit regression model. A questionnaire survey was carried out to collect the survey data; a

total of 356 samples took part in the study, and 310 were identified as valid samples. The results show that facilities, convenience, reliability, empathy, economics and timeliness have a significant impact on customer logistics service satisfaction, under the context of the growing number of Chinese Internet users and the expanding online retail market. Among them, facilities, economics and convenience are the most important factors. Therefore, it is recommended that logistics companies should improve their equipment or invest more into modern equipment to enable them to address the challenges for better quality service. In addition, they should also adjust the price of logistics services within the scope of cost control and provide customers with more diversified services to improve the convenience of customers when using logistics services.

Conflicts of Interest The authors declare that they have no conflicts of interest.

References

1. Guo, J., Xu, L., Xiao, G., Gong, Z.: Improving multilingual semantic interoperation in cross-organizational enterprise systems through concept disambiguation. IEEE Trans. Industr. Inf. **8**(3), 647–658 (2012)
2. Wu, J., Li, L., Xu, L.: A randomized pricing decision support system in electronic commerce. Decis. Support Syst. **58**, 43–52 (2014)
3. CNNIC: The 44th statistical report on internet development in China, Beijing. Available at: http://www.cac.gov.cn/2019-08/30/c_1124939590.htm, last accessed 2019/12/24
4. Christopher, M.: Logistics and supply chain management: strategies for reducing cost and improving service. Pitman Publishing, Financial Times, London (1998)
5. Mentzer, J.T., Flint, D.J., Hult, G.T.M.: Logistics service quality as a segment-customized process. J. Mark. **65**(4), 82–104 (2001)
6. Sudrajat, D., Amin, T.M.Z.: Strategic competitiveness improvement of indonesian logistics services companies through competitive advantage improvement: a conceptual model. Adv. Sci. Lett. **22**(12), 4540–4543 (2016)
7. Wang, X., Yuen, K.F., Wong, Y.D., Teo, C.C.: E-consumer adoption of innovative last-mile logistics services: a comparison of behavioural models. Total Qual. Manag. Bus. Excellence 1–27 (2018)
8. Thao, P.V., David, B.G., David, A.M.: Exploring logistics service quality in Hai Phong Vietnam. Asian J. Shipping Logistics (2019). https://doi.org/10.1016/j.ajsl.2019.12.001
9. Lin, C.C., Chang, C.H.: Evaluating skill requirement for logistics operation practitioners: based on the perceptions of logistics service providers and academics in Taiwan. Asian J. Shipping Logistics **34**(4), 328–336 (2018)
10. Limbourg, S., Giang, H.T.Q., Cools, M.: Logistics service quality: the case of Da Nang city. Procedia Eng. **142**, 124–130 (2016)
11. Lin, Y., Luo, J., Zhou, L., Ieromonachou, P., Huang, L., Cai, S. Ma, S.: The impacts of service quality and customer satisfaction in the e-commerce context. In: 2014 11th International Conference on Service Systems and Service Management (ICSSSM), June, 1–6 (2014). IEEE
12. Xu, C., et al.: Potential risk and its influencing factors for separated bicycle paths. Accid. Anal. Prev. **87**, 59–67 (2016)
13. Briggs, E., Landry, T.D., Daugherty, P.J.: Investigating the influence of velocity performance on satisfaction with third party logistics service. Ind. Mark. Manage. **39**(4), 640–649 (2010)

14. Rahmat, A.K., Faisol, N.: Manufacturers satisfaction on logistics service quality: operational, relational and national culture. Procedia Soc. Behav. Sci. **224**, 339–346 (2016)
15. Stopka, O., Černá, L., Zitrický, V.: Methodology for measuring the customer satisfaction with the logistics services. NAŠE MORE: znanstveno-stručni časopis za more i pomorstvo **63**(3), 189–194 (2016)
16. Tontini, G., Söilen, K.S., Zanchett, R.: Nonlinear antecedents of customer satisfaction and loyalty in third-party logistics services (3PL). Asia Pacific J. Mark. Logistics **29**(5), 1116–1135 (2017)
17. Duc N.L., Hong T.N., Phuc H. T.: Port logistics service quality and customer satisfaction: empirical evidence from vietnam. Asian J. Shipping Logistics (2019). https://doi.org/10.1016/j.ajsl.2019.10.003
18. Hua, W., Jing, Z.: An empirical study on e-commerce logistics service quality and customer satisfaction. WHICEB Proc. 269–275 (2015)
19. Jiang, K., Ling, F., Feng, Z., Wang, K., Shao, C.: Why do drivers continue driving while fatigued? an application of the theory of planned behaviour. Transp. Res. Part A Policy Pract. **98**, 141–149 (2017)
20. Li, X., Medal, H., Qu, X.: Connected heterogeneous infrastructure location design under additive service utilities. Transp. Res. Part B **120**, 99–124 (2019)
21. Ahmed, M.M., Franke, R., Ksaibati, et al.: Effects of truck traffic on crash injury severity on rural highways in wyoming using bayesian binary logit models. Accid. Anal. Prev. (2018)
22. Wang, S., Zhang, W., Qu, X.: Trial-and-error train fare design scheme for addressing boarding/alighting congestion at CBD stations. Transp. Res. Part B **118**, 318–335 (2018)

To Investigate the Hidden Gap between Traffic Flow Fundamental Diagrams and the Derived Microscopic Car Following Models: A Theoretical Analysis

Yang Yu, Jie Zhu, and Xiaobo Qu

Abstract Traffic flow fundamental diagram, or simply speed–density relationship and/or flow–density relationship, is the basis of traffic flow theories and road performance studies since it depicts the mathematical relationship among three traffic flow fundamental parameters—density, speed, and traffic flow. In this paper, through mathematical analyses and simulations, we find that for all existing fundamental diagram models, their derived microscopic car following models do not perform well and cannot reproduce the status of the stable flow described by the corresponding fundamental diagrams. The results indicate that there seems to exist a hidden gap between existing traffic flow fundamental diagrams and the corresponding microscopic car following models. We further discuss about the fundamental causes behind such gap and propose a simple yet incomplete solution at the end of this paper.

1   Introduction

Traffic flow fundamental diagram, or simply speed–density relationship and/or flow–density relationship, describes the fundamental relationship among traffic flow (vehs/h), speed (km/h), and traffic density (vehs/km) of a road segment. It is a very important part of macroscopic traffic theories and has been attracting a great many research attentions in the past decades [1–8]. Most models for describing speed–density relationships can be divided into either single-regime (use one curve to depict

Y. Yu (✉)
School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia
e-mail: Yang.Yu-4@student.uts.edu.au

J. Zhu · X. Qu
Department of Architecture and Civil Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden
e-mail: jie.zhu@chalmers.se

X. Qu
e-mail: xiaobo@chalmers.se

speed–density relationship in all traffic states) models such as Greenshields linear model [1], Underwood exponential model [4], and Greenberg logarithm model [2] or multiple-regime models (use different curves to depict speed–density relationships under different traffic states) such as Wu's four-state model [3].

On the other hand, microscopic car following models are designed to reproduce how vehicles are following one another on roadways. Most car following models are either formula-based (parametric model) [9–12] or machine-learning based (nonparametric model) [13–17], and they all take a vehicle's current status and surrounding traffic conditions as inputs and output the movement of the vehicle at next time step in the form of acceleration [10–12], speed [9], or moving distance [17].

Although numerous studies have focused on the development and optimization of macroscopic speed–density relationships (fundamental diagram models) and microscopic car following models, respectively, few studies took very intensive investigations into the potential relationships between them. In this paper, we fill this margin by investigating how microscopic car following models can be derived from macroscopic speed–density models (fundamental diagram models), as well as the features and defects of the derived car following models. We also find that there seems to exist a hidden gap between existing traffic flow fundamental diagrams and the corresponding microscopic car following models. We further discuss about the likelihood to bridge such gap and propose a simple yet incomplete solution.

The rest of the paper is organized as follows. Section 2 first introduces the general structure of traffic flow fundamental diagram models. Then, the corresponding general microscopic car following model is derived, which is followed by the analysis of the hidden gap between them. Section 3 discuss about the fundamental causes behind such gap and propose a simple yet incomplete solution. Section 4 concludes.

2   Theoretic Analysis

2.1   General Structure of Fundamental Diagram Models

Although traffic flow fundamental diagrams can be depicted by either single-regime or multiple-regime models, all these models in fact follow the same basic pattern: to describe traffic speed u as (a) function (s) of traffic density k so that traffic flow q can be depicted by a single fundamental traffic parameter—either u or k. Thus, a general structure for all existing fundamental diagram models can be described as follows:

$$u = f(k) \tag{1}$$

$$q = ku = kf(k) \tag{2}$$

where $f(k)$ can refer to any forms of function (single or piecewise function) that takes traffic density k as sole input and outputs traffic speed u. The above two equations (the fundamental diagram model) uniquely describe the stable status of a continuous traffic stream: the speeds of vehicles in the traffic stream tend to reach the value of $(u = f(k))$ when the density reaches the value of k. Meanwhile, the traffic flow tends to reach the value of $(q = kf(k))$. A typical speed–density relationship (Eq. (1)) and flow–density relationship (Eq. (2)) are illustrated in Fig. 1a, b, respectively.
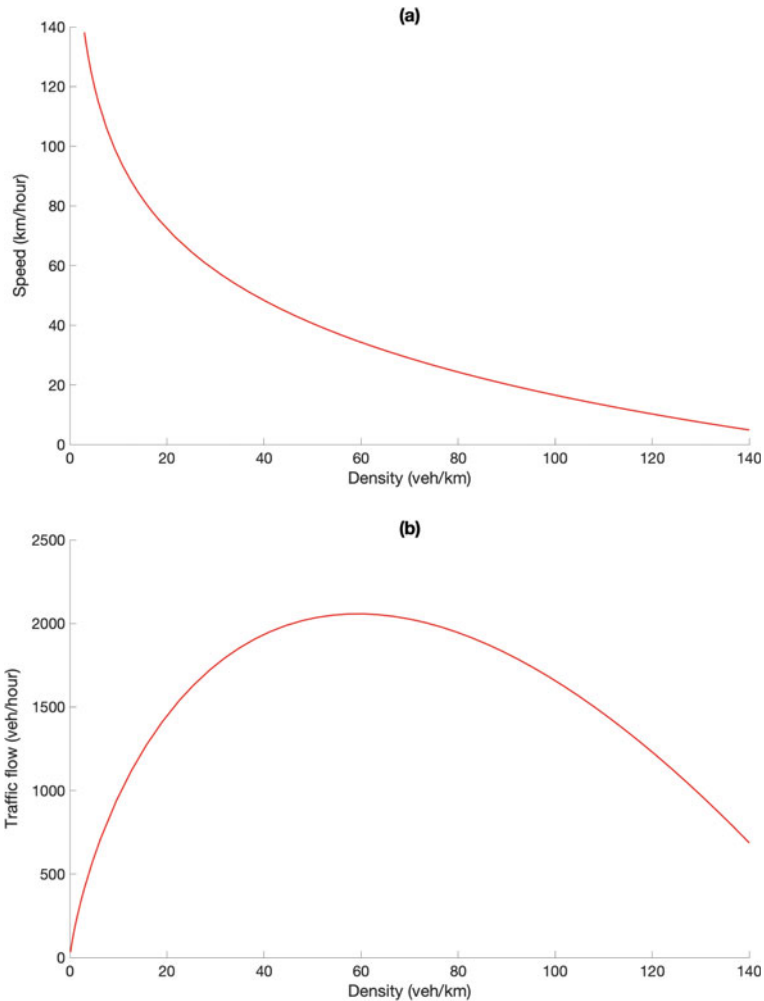
Fig. 1 Speed–density curve (**a**) and flow–density curve (**b**) of a Greenberg logarithm fundamental diagram model. Every point on the curves indicates a speed–density or flow-density value pairs that correspond to a stable flow state

2.2 Derivation of the Corresponding Microscopic Car Following Models

In order to derive the corresponding microscopic car following model from the fundamental diagram model, we can simply calculate the first-order derivative of Eq. (1) with regard to time t so that the left side term (traffic speed u) of Eq. (1) can result in vehicle acceleration a (the original unit of a generated is km/h^2, which can be easily transferred to the more common unit—m/s^2), which is a typical output of car following models since acceleration is directly linked to the control of engine power [14]. The above can be mathematically described as follows:

$$a \equiv \frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\mathrm{d}f(k)}{\mathrm{d}t} \tag{3}$$

Given that one can use the inverse of the spacing Δx between two successive vehicles to approximate the local density k, Eq. (3) can be further simplified as follows:

$$\begin{aligned}
a &= \frac{\mathrm{d}f(k)}{\mathrm{d}t} \\
&= \frac{\mathrm{d}f(k)}{\mathrm{d}k} \times \frac{\mathrm{d}k}{\mathrm{d}\Delta x} \times \frac{\mathrm{d}\Delta x}{\mathrm{d}t} \\
&= \frac{\mathrm{d}f(k)}{\mathrm{d}k} \times \frac{\mathrm{d}\frac{1}{\Delta x}}{\mathrm{d}\Delta x} \times \Delta v \\
&= \frac{\mathrm{d}f(k)}{\mathrm{d}k} \times \left(\frac{-1}{\Delta x^2}\right) \times \Delta v
\end{aligned} \tag{4}$$

Thus, the general form of microscopic car following models derived from any existing macroscopic speed–density relationships (fundamental diagram models) can be described as follows:

$$a = \lambda \Delta v \tag{5}$$

where Δv serves as the input stimulus and refers to the speed difference between a leading vehicle and its immediate follower, while λ refers to the sensitivity coefficient and is depicted by

$$\lambda = \left(\frac{-1}{\Delta x^2}\right) \frac{\mathrm{d}f(k)}{\mathrm{d}k} \tag{6}$$

2.3 Gap Between the Macroscopic and Corresponding Microscopic Traffic Models

In this subsection, we attempt to investigate the performance and defects of the fundamental diagram-based car following model. As can be easily observed from Eqs. (5) to (6), the microscopic car following model derived from any existing fundamental diagram model always holds the following features:

1. It takes use of the speed difference Δv as the only stimulus to predict the following vehicle's acceleration at the next time step;
2. The degree of the car following model's response to speed difference is determined by the sensitivity coefficient λ which increases/decreases with the increase/decrease of the spacing between the following vehicle and its leader;
3. The spacing Δx between the following vehicle and its leader does not directly act as an input stimulus for predicting the vehicle acceleration because it is not an individual term on the right-hand side of Eq. (5). Instead, it is only a part of the sensitivity coefficient for stimulus Δv, as was explained in Feature 2.

Therefore, it is not difficult to conclude that the above car following model would always force the vehicle it controls to reach a 'fake stable car following status' as long as the speed difference Δv reaches 0 for the first time: the following vehicle would never change its speed again (always keep the same speed as its immediate leader) regardless of the actual spacing between the vehicle and its leader, and this 'fake stable status' would gradually pass from the first follower in a platoon to the last follower until the entire platoon moves with the same speed but with different spacing (different local densities) between one another. In other words, the car following model cannot reproduce the status of the stable flow described by the corresponding fundamental diagram model which assumes that the same, stable speed of vehicles in a platoon would correspond to the same spacing between two successive vehicles (the stable local density), as was explained in Fig. 1. On top of that, the above car following model could lead to extremely dangerous situations since it may force the vehicle to always follow its immediate leader with the same speed but at a very small spacing. On the other hand, it could be very unreasonable since the model may also force the vehicle to keep the same speed as its leader given that the spacing is still very large. To summarize, there seems to exist a hidden gap between existing traffic flow fundamental diagrams and the corresponding microscopic car following models.

Now, we take use of the famous Greenberg logarithm speed–density model (fundamental diagram model) to help illustrate the above findings. By replacing the general function $f(k)$ in Eq. (1) with Greenberg model function, we have

$$u = u_o \ln \frac{k_j}{k} \tag{7}$$

where u_o is the optimal speed of vehicles in a platoon at the capacity point and k_j is the jam density. By approximating the independent variable k using the inverse

Table 1 Calibrated results of two model parameters

	u_o	k_j
Calibrated value	34.7 km/h	161 veh/km

of spacing $\frac{1}{\Delta x}$ and adopting the same process as was in Sect. 2.2 to calculate the first-order derivative of Eq. (7) with regard to time t, we have

$$a = \frac{u_o}{\Delta x} \Delta v \tag{8}$$

which happens to be a special form of the general motors car following model [18] that was developed decades ago.

To be consistent with previous studies regarding fundamental diagram [7, 8, 19], we calibrated both Eqs. (7) and (8) using the 47,185 observations of speed–density data collected from Georgia State Route 400. The calibrated values for the two parameters (u_o and k_j) are listed in Table 1.

Now, we try to set up a reasonable testing environment: without loss of genuinity, we create a platoon of 20 vehicles following one another, and the initial spacing between two successive vehicles is randomly generated from a normal distribution with mean value equals to the inverse of optimal density k_o at capacity point (k_o can be easily calculated from Eq. (7) given that u_o and k_j are already known). Similarly, the initial speed of each following vehicle is also randomly generated from a normal distribution with mean value equals to the calibrated optimal speed u_o. At last, the speed of the first vehicle in the platoon is set fixed as the optimal speed u_o.

By using the calibrated car following model (Eq. 8) derived from Greenberg fundamental diagram model (Eq. 7) to control all the 19 followers in the above vehicle platoon, simulation results (Fig. 2) show that upon the platoon reaches its 'stable status' (when all vehicles reach and keep the same speed as the first vehicle, which is the optimal speed u_o), the spacing between two successive vehicles varies largely (from only 6.73 to 30.16 m with a mean value of about 13 m) rather than reaching close to the same, stable spacing (17 m, which corresponds to the inverse of optimal density k_o). The results perfectly validated our aforementioned deduction that the fundamental diagram-derived car following model cannot reproduce the status of the stable flow described by the corresponding fundamental diagram.

In addition, we repeat the above simulation by only changing the speed profile of the first vehicle from uniform motion to a decelerated motion followed by an accelerated motion to further test whether the above car following model would easily lead to dangerous situations. The simulation results are displayed in Fig. 3, from which we can easily observe that rear-end collisions (definitely a very dangerous situation) have easily occurred when the leading vehicle drives in a more random (natural) way, just as we have analysed before.
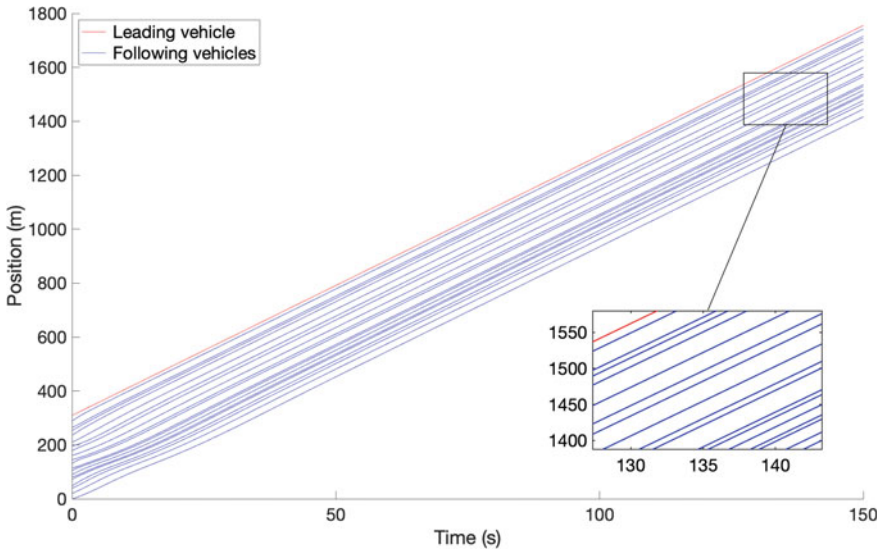
Fig. 2 Simulation results of a platoon of vehicles following a fixed-speed leader
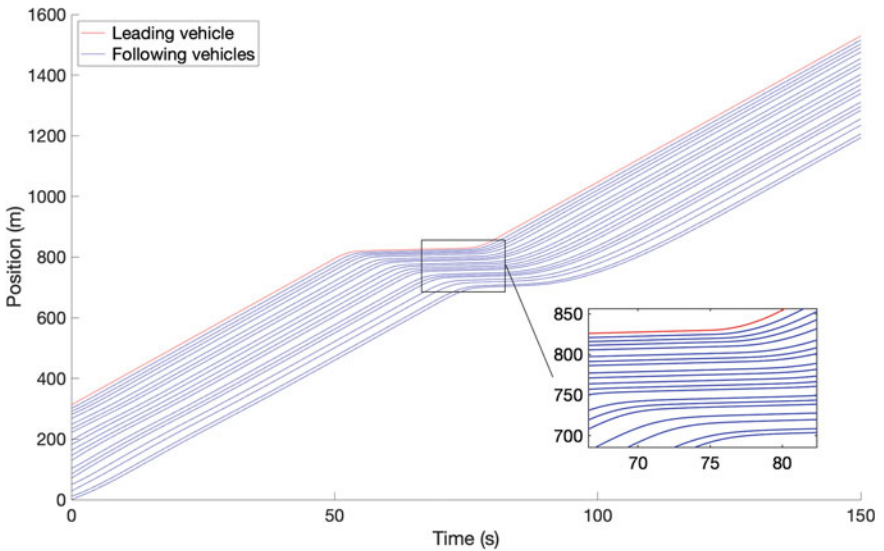

Fig. 3 Simulation results of a platoon of vehicles following a speed-changing leader. *Note* since the spacing between two vehicles should include the vehicle length of around 5 metres, the actual net spacing between two vehicles in the magnified zone is already very close to or even smaller than 0, which indicates the situation of rear-end collisions

3   Further Discussions

In this section, we attempt to discuss more about the fundamental cause behind the aforementioned gap between the existing fundamental diagram models and their derived microscopic car following models, and the likelihood to address such gap.

As was analysed in Sect. 2.3, the 'fake stable car following status' would be easily yielded by any car following models derived from any of the existing fundamental diagram models. This is caused by the fact that these derived car following models take speed difference Δv as the only input stimulus but fail to separately consider the impact of spacing Δx between two vehicles, which, however, is another very important factor every driver will consider during driving and another input stimulus used by most car following models such as models of optimal velocity model family [10, 12, 20–22], models of intelligent driver model family [11, 23, 24], and models of the machine learning-based model family [14–17]. But if we look back to Eq. (4) which showed the detailed process for calculating the first-order derivative of general macroscopic speed–density relationship (general fundamental diagram model), we can find that the only input stimulus Δv of derived car following model in fact comes from the only variable k on the right-hand side of the general speed–density relationship through replacing density k using the inverse of spacing $\frac{1}{\Delta x}$ and then calculate its derivative with regard to time t. In other words, the aforementioned defects of the derived car following models (also the gap between the macroscopic traffic model and its derived microscopic model) will always exist for all kinds of speed–density models (fundamental diagram models). The only possible solution to nicely address such gap is to modify the structure of speed–density relationships by adding more variable (s) to the right-hand side of the model, which, however, would greatly improve the model complexity and is also impractical at present since all the three fundamental traffic parameters (speed, density, and traffic flow) are already nicely integrated into the existing fundamental diagram models.

Although changing of the general structure of the widely accepted speed–density models is difficult and impractical, one can still address the above defects of all the fundamental diagram-derived car following models by simply introducing the difference between desired stable spacing Δx_{stable} corresponding to the current speed[1] and existing spacing Δx as a second stimulus of the car following models, which would enable the models to keep re-adjusting the accelerations/decelerations of the following vehicle even when the speed difference Δv is 0 so that stable car following platoon would only be formed when every vehicle both reaches the same speed as the first vehicle and keeps the desired spacing (corresponds to the inverse of the stable density at that speed). However, by modifying the above car following models like this, the strict derivation relationship between the macroscopic traffic models and their corresponding microscopic models would no longer hold.

---

[1]Δx_{stable} can be easily acquired from the corresponding macroscopic speed–density relationship: simply use current speed u of the following vehicle as input to output the corresponding stable density k, which is $\frac{1}{\Delta x_{\text{stable}}}$.

4 Conclusion

Although numerous studies have focused on the development/optimization of macroscopic speed–density relationships (fundamental diagram models) and microscopic car following models, respectively, few studies took intensive investigations into the potential relationships between them. In this paper, we fill the above margin by investigating how microscopic car following models can be derived from macroscopic speed–density models (fundamental diagram models), as well as the features and defects of the derived car following models. Based on the theoretical analyses and simulation results, we point out that for all existing speed–density models (fundamental diagram models), their derived microscopic car following models would always take speed difference as the only stimulus to produce next-step accelerations, which is likely to lead to safety issues such as rear-end collisions and also cannot reproduce the status of the stable flow described by the corresponding fundamental diagrams. In other words, there seems to exist a hidden gap between existing traffic flow fundamental diagrams and the corresponding microscopic car following models. At the end of this paper, we further discuss about the likelihood to bridge such gap and propose a simple yet incomplete solution.

References

1. Greenshields, B., Channing, W., Miller, H.: A study of traffic capacity. In: Highway Research Board Proceedings, 1935, vol. 1935. National Research Council (USA), Highway Research Board
2. Greenberg, H.: An analysis of traffic flow. Oper. Res. **7**(1), 79–85 (1959)
3. Wu, N.: A new approach for modeling of fundamental diagrams. Transp. Res. Part A: Policy Pract. **36**(10), 867–884 (2002)
4. Underwood, R.T.: Speed, Volume, and Density Relationship: Quality and Theory of Traffic Flow, (1961), pp. 141–188. Yale Bureau of Highway Traffic, New Haven, Connecticut (2008)
5. Wang, H., Li, J., Chen, Q.-Y., Ni, D.: Logistic modeling of the equilibrium speed–density relationship. Transp. Res. Part A: Policy Pract. **45**(6), 554–566 (2011)
6. Wang, H., Ni, D., Chen, Q.Y., Li, J.: Stochastic modeling of the equilibrium speed–density relationship. J. Adv. Transp. **47**(1), 126–150 (2013)
7. Qu, X., Wang, S., Zhang, J.: On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models. Transp. Res. Part B: Methodol. **73**, 91–102 (2015)
8. Qu, X., Zhang, J., Wang, S.: On the stochastic fundamental diagram for freeway traffic: model development, analytical properties, validation, and extensive applications. Transp. Res. Part B: Methodol. **104**, 256–271 (2017)
9. Gipps, P.G.: A behavioural car-following model for computer simulation. Transp. Res. Part B: Methodol. **15**(2), 105–111 (1981)
10. Bando, M., Hasebe, K., Nakayama, A., Shibata, A., Sugiyama, Y.: Dynamical model of traffic congestion and numerical simulation. Phys. Rev. E **51**(2), 1035 (1995)
11. Treiber, M., Hennecke, A., Helbing, D.: Congested traffic states in empirical observations and microscopic simulations. Phy. Rev. E. **62**(2), 1805–1824 (2000). Available: https://link.aps.org/doi/10.1103/PhysRevE.62.1805
12. Jiang, R., Wu, Q., Zhu, Z.: Full velocity difference model for a car-following theory. Phys. Rev. E **64**(1), 017101 (2001)

13. Panwai, S., Dia, H.: Neural agent car-following models. IEEE Trans. Intell. Transp. Syst. **8**(1), 60–70 (2007)
14. Zhou, M., Qu, X., Li, X.: A recurrent neural network based microscopic car following model to predict traffic oscillation. Transp. Res. Part C: Emerg. Technol **84**, 245–264 (2017)
15. Xu, C., et al.: Potential risk and its influencing factors for separated bicycle paths. Accid. Anal. Prev. **87**, 59–67 (2016)
16. Qu, X., Yu, Y., Zhou, M., Lin, C.-T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach. Appl. Energy **257**, 114030 (2020)
17. He, Z., Zheng, L., Guan, W.: A simple nonparametric car-following model driven by field data. Transp. Res. Part B: Methodol. **80**, 185–201 (2015)
18. Chakroborty, P., Kikuchi, S.: Evaluation of the general motors based car-following models and a proposed fuzzy inference model. Transp. Res. Part C: Emerg. Technol. **7**(4), 209–235 (1999)
19. Zhang, J., Qu, X., Wang, S.: Reproducible generation of experimental data sample for calibrating traffic flow fundamental diagram. Transp. Res. Part A: Policy Pract. **111**, 41–52 (2018)
20. Helbing, D., Tilch, B.: Generalized force model of traffic dynamics. Phys. Rev. E **58**(1), 133 (1998)
21. Jin, S., Wang, D., Tao, P., Li, P.: Non-lane-based full velocity difference car following model. Phy. A **389**(21), 4654–4662 (2010)
22. Yu, Y., Jiang, R., Qu, X.: A modified full velocity difference model with acceleration and deceleration confinement: Calibrations, validations, and scenario analyses. IEEE Intell. Transp. Syst. Maga. (2019)
23. Kesting, A., Treiber, M., Helbing, D.: Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. Philos. Trans. Roy. Soc. London A: Math. Phy. Eng. Sci. **368**(1928), 4585–4605 (2010)
24. Zhou, M., Qu, X., Jin, S.: On the impact of cooperative autonomous vehicles in improving freeway merging: a modified intelligent driver model-based approach. IEEE Trans. Intell. Transp. Syst. **18**(6), 1422–1428 (2017)

A Leader-Based Vehicle Platoon Control Strategy at Signalized Intersections Considering Efficiency

Jian Zhang, Tie-Qiao Tang, Yang Yu, and Xiaobo Qu

Abstract In this paper, we propose a leader-based control strategy for vehicle platoon at signalized intersections. The speed guidance which regards the states of followers and the signal phase and timing (SPaT) information is adopted for the leader of the platoon. The Cooperative Adaptive Cruise Control (CACC) with a piecewise policy is used to control the followers. The efficiency of the signalized intersection is specifically considered when constructing the control strategy. The simulation results show that the proposed strategy can control the vehicle platoon smoothly cross signalized intersections without stop and ensure the traffic efficiency simultaneously.

1 Introduction

The transportation system is developing toward automation and intelligence owing to the emergence of sensor and vehicular network technology [1, 2]. Traffic platooning is an important method that can substantially increase road capacity, energy economy, and safety in the smart transportation system and thus attracts much attention from researchers.

A vehicle platoon is defined as a group of vehicles that travel together with nearly the same speed and small headway. In order to maximize the advantages of vehicle platoon driving, a large body of research has studied the protocols for platoon maneuvers on highways from the platoon level, including platoon formation, split, merge,

J. Zhang · T.-Q. Tang (✉)
School of Transportation Science and Engineering, Beihang University, 100191 Beijing, China
e-mail: tieqiaotang@buaa.edu.cn

Y. Yu
School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia

X. Qu
Department of Architecture and Civil Engineering, Chalmers University of Technology, SE412 96 Gothenburg, Sweden

99

and exit [1, 3]. As for the vehicle control level, another strand of research studied the Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC) that can help a vehicle track its preceding vehicle by using real-time state information to achieve the platooning driving [4, 5]. However, these control strategies cannot be directly used when a vehicle platoon is traveling along a signalized road since they do not explicitly consider the SPaT information. Signalized intersection is a crucial point to produce traffic delays, extra safety hazards, and excessive energy consumption. Various advanced traffic control strategies, such as speed guidance, have been developed to improve traffic performance measures on signalized arterial [6, 7].

In this paper, we propose a leader-based control method which combines speed guidance strategy and CACC for vehicle platoon when passing through a series of signalized intersections. Moreover, we explicitly consider the efficiency of the signalized intersection, i.e., the capacity of the signalized intersection, when constructing the control strategy. The rest of the paper is organized as follows: The control strategy is constructed in Sect. 2, simulations are conducted to validate the proposed strategy in Sect. 3, and some conclusions are summarized in Sect. 4.

2   Control Strategy

It is assumed that there is a predetermined platoon that consists of two parts, the leader and the followers, passing through a one-lane signalized artery where overtaking is forbidden. The leader should take responsibility for the platoon, is a manual driving vehicle equipped with a speed guidance system, and the followers are connected vehicles controlled by CACC. Both the speed guidance and CACC are easy to be achieved in the state-of-art without a need for expensive sensors widely used in automatic vehicles, which make the method easy to promote at the moment.

First, we introduce the CACC system equipped on the followers. According to the literature, there are two popular policies of CACC, the constant spacing (CS) policy [8] and constant time-gap (CT) policy [9]. The CT policy is more favorable among the studies of CACC on highways without interruptions, and some filed tests also have proven its superior performance. However, CT policy is not suited for this study since when vehicles pass through intersections, a low-speed situation or stop and start situation cannot be eliminated. It is obvious that if the speed is very low, the follower cannot keep a constant time-gap, as illustrated in Fig. 1 denoted by the slanted solid line and slanted dashed line, the spacing approaches zero as speed approaches zero. From the previous study on manual driving features, we can also find that one prefers to keep a relatively constant time-gap at high speed while a relatively constant spacing at low speed. As a consequence, a piecewise CACC policy is adopted in this study with the consideration of different speed situations. The policy can be formulated as follows:
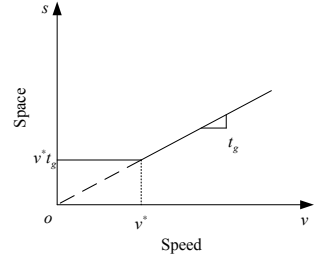
$$s(v) = \begin{cases} vt_g & \text{if } v > v \\ v * t_g & \text{else} \end{cases} \tag{1}$$

where $s(v)$ is the inter-vehicle space, t_g is the constant time-gap (in the range of 0.6–1.1 s based on the previous studies), v^* is the speed boundary value of CT policy and CS policy. This is illustrated in Fig. 1 denoted by the slanted solid line and horizontal solid line.

The technical details of the method are out of the purpose of this paper, so we only propose the mathematical model and test the performance of this method through simulations. According to Milanés and Shladover [3], the car-following behavior of CACC controlled vehicles can be well simulated by the following model:

$$e_n(t + \Delta t) = x_{n-1}(t) - x_n(t) - s(v_n(t)), \tag{2}$$

$$v_n(t + \Delta t) = v_n(t) + k_1 e_n(t + \Delta t) + k_2 \dot{e}_n(t + \Delta t). \tag{3}$$

where $x_n(t)$ and $v_n(t)$ are, respectively, the position, the velocity of the nth vehicle at time t, e_n is the spacing error, which is the deviation of the expected space headway s and the real one. Δt is the time step, k_1 and k_2 are two control parameters to decrease the spacing error and determine the velocity of the next time step, set to be $k_1 = 0.45$ and $k_2 = 0.25$ based on the experimental data [3]. Here, we should note that bounded acceleration is applied in the simulation according to the real vehicle's limitations.

Next, we introduce the dynamic of the leader in the platoon. Since we have assumed that the leader is a manual driving vehicle with a speed guidance system, the dynamic model is also divided into two parts. For the manual driving behavior, the simplified Gipps model [10] is implemented. A modified car-following model with speed suggestion is used to describe the driving behavior with speed guidance. The acceleration calculator of Gipps model is divided into two parts, the free-flow acceleration a_n^f and the congested acceleration a_n^c, which can be formulated as follows:

$$a_n^f(t) = 2.5 \cdot a_{\max} \cdot \left(1 - \frac{v_n(t)}{v_{\max}}\right) \cdot \sqrt{0.025 + \frac{v_n(t)}{v_{\max}}}, \tag{4}$$

$$a_n^c(t) = \frac{1}{T} \cdot \left[\frac{1}{\tau} \left(x_{n-1}(t) - x_n(t) - s_j + \frac{v_n(t)^2 - v_{n-1}(t)^2}{2 \cdot a_{\min}} \right) - v_n(t) \right]. \quad (5)$$

where $a_{\max}$ is the maximum acceleration, $a_{\min}$ is the maximum deceleration, which is negative, $v_{\max}$ is the maximum speed in free flow, T is the sensitivity coefficient, τ is the reaction time of drivers, s_j is the jam spacing between cars. Here, we should note that Eq. (5) does not consider the influence of traffic light, if the signal is green, we can use it to calculate congested acceleration directly, but if the signal is red, the congested acceleration should be calculated as follows:

$$a_n^c(t) = \min \left\{ \begin{array}{l} \frac{1}{T} \cdot \left[\frac{1}{\tau} \left(x_{n-1}(t) - x_n(t) - s_j + \frac{v_n(t)^2 - v_{n-1}(t)^2}{2 \cdot a_{\min}} \right) - v_n(t) \right], \\ \frac{1}{T} \cdot \left[\frac{1}{\tau} \left(\text{LT} - x_n(t) - s_j + \frac{v_n(t)^2}{2 \cdot a_{\min}} \right) - v_n(t) \right] \end{array} \right\}, \quad (6)$$

where LT is the position of the stop line.

In the simulation, we assume that the acceleration is constant during each time step, and the speed cannot be negative or larger than the speed limit, with the bounded acceleration assumption, the acceleration a_n can be calculated as follows:

$$a_n(t) = \max \left\{ a_{\min}, \frac{-v_n(t)}{\Delta t}, \min \left\{ a_n^f(t), a_n^c(t), \frac{v_{\max} - v_n(t)}{\Delta t} \right\} \right\}. \quad (7)$$

When the leader vehicle gets the guidance speed v_g from the speed guidance system, the driving behavior model is modified by replacing $v_{\max}$ by v_g in Eq. (4) according to [11].

When the platoon is approaching the signalized intersection, the speed guidance system will collect the speed and position of each vehicle through V2V communication as well as signal information through V2I communication. We assume that the speed and position data of other vehicles which are not equipped with communication systems can be collected by the loop detectors installed at the upstream of the intersection and the entrance of the intersection. Then, the expected arrival time at the intersection of each vehicle T_n^a can be estimated based on the collected data:

$$T_n^a = \frac{\text{LT} - x_n(t)}{v_n(t)}. \quad (8)$$

Once T_n^a is estimated, we can determine if the vehicle is encountering a red signal light when it reaches the intersection based on the signal timing information. If vehicle n will encounter a red light, we should discuss this situation according to the vehicle's type. If vehicle n is a follower of the platoon, we cannot adjust its motion directly since it is controlled by the CACC system. However, the driving behavior of the leader can be adjusted to let the follower pass through the intersection without a stop, thus increasing the capacity of the intersection. If vehicle n is a leader of the platoon, we can adjust its speed via the speed guidance system, but the adjustment should be made with the consideration of followers since the leader should take the
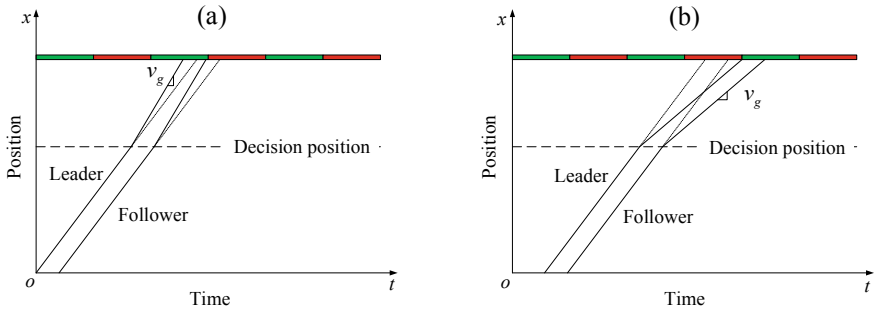
Fig. 2 Situations when the leader adjusts the velocity, where **a** is acceleration and **b** is deceleration

responsibility of the vehicle platoon and let all vehicles pass through the intersection at the same green phase.

There are two situations in which the leader adjusts the velocity, see Fig. 2, one is to accelerate and pass the intersection at the current green phase, while the other is deceleration and goes through the intersection at the next green phase.

For the acceleration situation, the leader should keep the last follower arrive at the stop line before the red light on, and the v_g should not be greater than $v_{\max}$. For the deceleration situation, the leader should arrive at the stop line after the green light on, and the last follower should pass through the intersection in the same green phase.

With the consideration that the vehicle's acceleration/deceleration is limited, the realistic trajectories of the vehicle platoon are different from those illustrated in Fig. 2. The guidance speed can be calculated based on a constant acceleration/deceleration assumption:

$$
v_g = \begin{cases} a_g\left[\left(\dfrac{v_0}{a_g}+t_e\right)-\sqrt{\left(\dfrac{v_0}{a_g}+t_e\right)^2-\dfrac{2}{a_g}\left(\dfrac{v_0^2}{2}+s\right)}\right] \\ d_g\left[\left(\dfrac{v_0}{d_g}-t_e\right)+\sqrt{\left(\dfrac{v_0}{d_g}-t_e\right)^2-\dfrac{2}{d_g}\left(\dfrac{v^2}{2}-s\right)}\right] \end{cases}, \tag{9}
$$

$$
s = LT - x_n, \tag{10}
$$

$$
t_e = \begin{cases} t_{\text{ge}} - t_d - k \cdot t_g \\ t_{\text{re}} - t_d \end{cases}. \tag{11}
$$

where a_g and d_g are the constant acceleration and deceleration which are positive, t_e is the expected time that the leader travels from the decision position to the intersection, t_d is the current time at decision position, t_{ge} is the green end time of the current signal phase, k is the number of followers in the platoon. It is obvious that green time should

be larger than $k \cdot t_g$. Otherwise, it is not possible for all vehicles in the platoon to pass through the intersection in the same green phase. t_{re} is the red end time of the current signal phase. Notice that if $v_g \geq v^*$, the followers will obey the CT policy to go through the intersection, but if $v_g < v^*$, the control policy of the CACC system will become CS. As illustrated in Fig. 3a, the time headway between the adjacent vehicles in the platoon will be larger than t_g, which leading to a decrease in intersection capacity. Therefore, a remedy should be taken to ensure the efficiency of the intersection. The main idea of the remedy is to ensure the platoon speed when passing through the intersection is not less than v^*.

As shown in Fig. 3b, the leader should slow down first and then speed up to a speed which is not less than v^* to pass through the intersection. For simplicity, we assume that the end speed is equal to v^*. The speed trajectory can be illustrated in Fig. 4, the vehicle will first decelerate from v_0 to v_g in time t_1, then it will travel at a constant speed v_g for time $(t_2 - t_1)$, and after that, it will accelerate to v^* and pass through the intersection. Therefore, the time headway between the adjacent vehicles when crossing the intersection will always be equal to t_g, which guarantees the capacity of the intersection further. To achieve the speed guidance strategy, v_g, t_1 and t_2 are concerned and can be calculated as follows:

$$v_g = \frac{a_g d_g}{a_g + d_g} \left[\xi + \sqrt{\xi^2 - \frac{a_g + d_g}{\left(a_g d_g\right)^2} \left(a_g v_0^2 + d_g v *^2 - 2 a_g d_g s \right)} \right], \qquad (12)$$
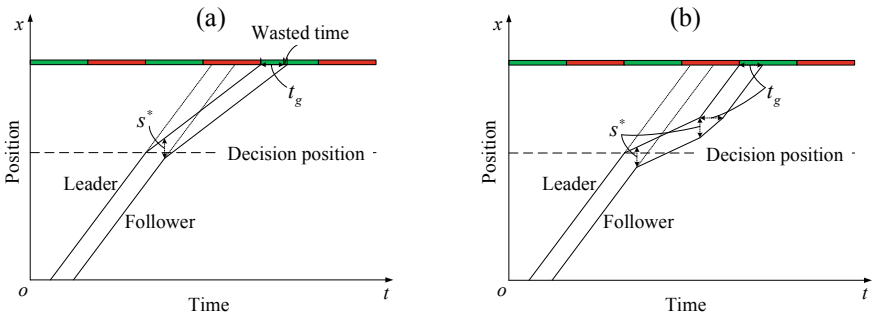
Fig. 3 **a** Decrease in intersection capacity and **b** the remedy trajectory

Fig. 4 Speed trajectory of the deceleration-and-acceleration strategy
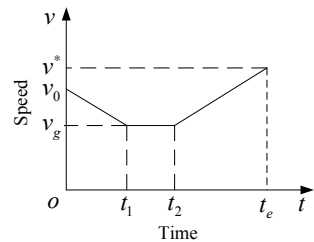
$$\xi = \frac{a_g v_0 + d_g v^*}{a_g d_g} - t_e, \tag{13}$$

$$t_e = t_{\text{re}} - t_d, \tag{14}$$

$$t_1 = \frac{v_0 - v_g}{d_g}, \tag{15}$$

$$t_2 = t_e - \frac{v^* - v_g}{a_g}. \tag{16}$$

3 Numerical Experiments

In this section, we conduct some numerical experiments to investigate the performance of the proposed control strategy. To validate the proposed control strategy, we conduct a comparison between manual and controlled vehicle platoons. The high-precision GPS device collects the driving data of a four-vehicle-platoon at the Vehicle Network and Intelligent Vehicle Test Site of Chang'an University in Xi'an, China. The test road is a 2.4 km long high-speed road with four signal lights. Note the test road is two-lane, but only the inner lane of the ring road was used for data collection. Overtaking was not allowed during the collection of driving data.

The simulation scenario is the same as the data collection scenario. The time step in the simulations is 0.1 s. The parameters of the Gipps model are as follows: $a_{\max} = 3$ m/s^2, $a_{\min} = -4$ m/s^2, $v_{\max} = 11.1$ m/s, $T = 1.2$ s, $\tau = 1.6$ s, $s_j = 10$ m. The comparison result is shown in Fig. 6, where the solid black line denotes the human driving data, and the blue dashed line denotes the simulation data. From Fig. 5, we can find that the manual platoon will slow down or stop at signalized intersections since it cannot obtain the SPaT information in advance, whereas the vehicle platoon

Fig. 5 Trajectories of manual vehicle platoon (solid black line) and controlled vehicle platoon (dashed blue line)
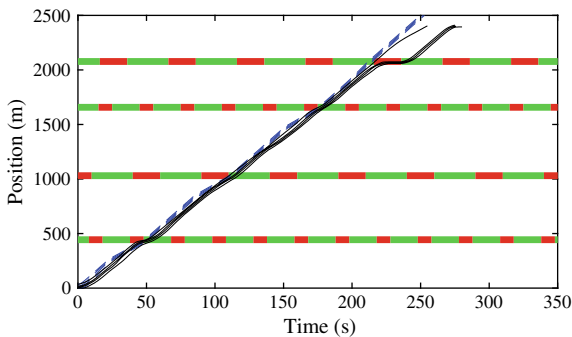
controlled by the proposed strategy can pass all intersections smoothly. The trajectories also suggest that the proposed control strategy can enhance traffic efficiency. Due to the cumulative effect of deceleration at the first three intersections, the manual vehicle platoon cannot pass the fourth intersection without stopping. However, the controlled vehicle platoon can always pass the intersection with a relatively high speed, which guarantees traffic efficiency.

Figure 6 shows the platoon length of the manual vehicle platoon and controlled vehicle platoon. From Fig. 6, we can find that the platoon length of the human driving vehicle platoon is greater than that of the vehicle platoon controlled by the proposed strategy in most instances. The range of controlled platoon length is much smaller than that of the human driving vehicle platoon, which indicates the stability of the controlled vehicle platoon. Figure 7a, b shows the space headway of followers in the manual vehicle platoon and controlled vehicle platoon, respectively. Note that the ordinate ranges of the two figures are different. The results of Fig. 7 show that the proposed control strategy can help vehicles in the platoon keep a relatively small and stable space, which can improve traffic efficiency and safety.
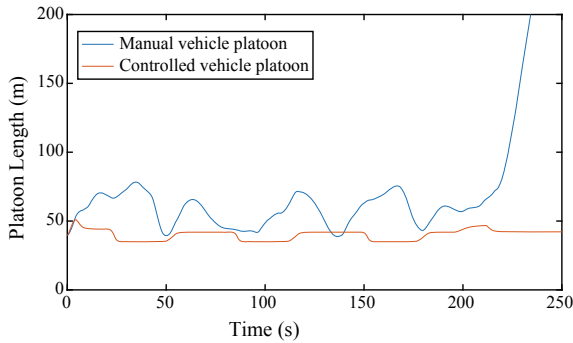

Fig. 6 Platoon length of manual vehicle platoon and controlled vehicle platoon
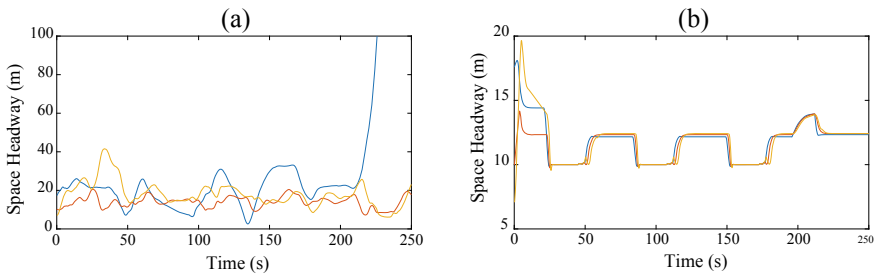

Fig. 7 Space headway of followers in the vehicle platoon, where **a** is the manual vehicle platoon, and **b** is the controlled vehicle platoon

4   Conclusion

In recent years, the transportation system is developing toward automation and intelligence. Vehicle platooning is an essential means to increase road capacity, energy economy, and safety in the smart transportation system. In this paper, we propose a leader-based vehicle platoon control strategy at signalized intersections. The speed guidance with consideration of following vehicles is adopted to give speed advice for the leader of the platoon. A piecewise CACC is used to control the followers. We individually consider traffic efficiency when designing the control strategy. A comparison of a manual vehicle platoon and a controlled vehicle platoon is conducted. The simulation results show that the proposed control strategy can help the vehicle platoon smoothly pass the signalized intersection and increase traffic efficiency.

However, we only propose the simulation model of the control strategy without the technical details. Moreover, the energy consumption is not studied in the present paper. Given the limitations of this paper, we should investigate the comprehensive control problem of the vehicle platoon and propose a control model that can be validated by some macroscopic methods, such as the stochastic fundamental diagram [12, 13]. With the electrification of the transport sector [14], further research should also include the control strategy of the electric vehicle platoon.

References

1. Zhou, M., Qu, X., Jin, S.: On the impact of cooperative autonomous vehicles in improving freeway merging: A modified intelligent driver model based approach. IEEE Trans. Intell. Transp. Syst. **18**, 1422–1428 (2017)
2. Li, X., Ghiasi, A., Xu, Z., Qu, X.: A piecewise trajectory optimization model for connected automated vehicles: Exact optimization algorithm and queue propagation analysis. Transp. Res. Part B **118**, 429–456 (2018)
3. Duret, A., Wang, M., Ladino, A.: A hierarchical approach for splitting truck platoons near network discontinuities. Transp. Res. Part B (2019)
4. Xu, C., Yang, Y., Jin, S., Qu, Z., Hou, L.: Potential risk and its influencing factors for separated bicycle paths. Accid. Anal. Prev. *87*, 59–67 (2016). https://doi.org/10.1016/j.aap.2015.11.014
5. Flores, C., Milanés, V.: Fractional-order-based ACC/CACC algorithm for improving string stability. Transp. Res. Part C **95**, 381–393 (2018)
6. Zhou, M., Yu, Y., Qu, X.: Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: A reinforcement learning approach. IEEE Trans. Intell. Transp. Syst. **21**, 433–443 (2020)
7. Bie, Y., Xiong, X., Yan, Y., Qu, X.: Dynamic headway control for high-frequency bus line based on speed guidance and intersection signal adjustment. Comput Aided Civ. Infrastruct. Eng. **35**, 4–25 (2020)
8. Darbha, S., Rajagopal, K.R.: Intelligent cruise control systems and traffic flow stability. Transp. Res. Part C **7**, 329–352 (1999)

9. Rajamani, R., Shladover, S.E.: Experimental comparative study of autonomous and co-operative vehicle-follower control systems. Transp. Res. Part C **9**, 15–31 (2001)
10. Gipps, P.G.: A behavioural car-following model for computer simulation. Transp. Res. Part B **15**, 105–111 (1981)
11. Ubiergo, G.A., Jin, W.L.: Mobility and environment improvement of signalized networks through vehicle-to-infrastructure (V2I) communications. Transp. Res. Part C **68**, 70–82 (2016)
12. Qu, X., Wang, S., Zhang, J.: On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models. Transp. Res. Part B **73**, 91–102 (2015)
13. Qu, X., Zhang, J., Wang, S.: On the stochastic fundamental diagram for freeway traffic: Model development, analytical properties, validation, and extensive applications. Transp. Res. Part B **104**, 256–271 (2017)
14. Qu, X., Yu, Y., Zhou, M., Lin, C.T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach. Appl. Energy **257**, 114030 (2020)

The Impact of Increasing Minor Arterial Flow on Arterial Coordination: An Analysis Based on MAXBAND Model

Liang Xu, Lixiao Shen, and Xiaobo Qu

Abstract With the progress of urbanization, car ownership is experiencing explosive growth in China, which leads to heavy pressure on the urban road network. Arterial coordination strategy has been proved an effective method to avoid or alleviate traffic congestion. However, with the increasing proportion of flow on the minor arterial, arterial coordination efficiency might be affected. To figure out the problem, a numerical test is conducted by designing eight scenarios with different proportion of through movement and left turn flow on the minor arterials. MAXBAND model is applied for optimizing signal plans. The results show that average delay for vehicles on the arterials increases with the increasing of proportion of through movement flow, as well as the entire average delay. Average delay for vehicles on the minor arterials and two-way bandwidth decreases at same time. In other words, when the proportion of minor arterial flow increases, the arterial coordination efficiency would be reduced, especially for increasing left turn flow. This work reveals the improvement direction for arterial coordination.

Keywords Arterial coordination · MAXBAND model · Coordination efficiency

1   Introduction

With the rapid economic development and expanding the size of cities, car ownership is experiencing explosive growth in China, which has reached 217 million by the

L. Xu (✉)
College of Civil Engineering and Architecture, Zhejiang University, 310058 Hangzhou, Zhejiang, China
e-mail: xuliangccea@zju.edu.cn

L. Shen
The Architectural Design and Research Institute of Zhejiang University Co. Ltd, 310058 Hangzhou, Zhejiang, China

X. Qu
Department of Architecture and Civil Engineering, Chalmers University of Technology, 41258 Gothenburg, Sweden

end of 2017 according to the report published by traffic administration bureau of the Ministry of Public Security. The rapid increase in car ownership leads to heavy pressure on the urban road network, which makes traffic congestion and accident happen more frequently [1, 2]. Focused on these troubles, various methods were developed to enhance traffic efficiency, including developing efficient driving strategy for automated vehicles [3, 4], encouraging residents to travel "green" [5–7], and so on. Among them, traffic signal control strategy has been proved to be an effective method to avoid or alleviate traffic congestion. It is reported that appropriate control strategy can reduce the average vehicle delay and improve road capacity based on statistic data from America, Germany, Japan, and other European countries [8].

Up to now, research on traffic signal control strategy can be roughly classified into two categories, isolated control and coordination control. With increased car traffic and extended road network, the connection between adjacent intersections becomes closer. The performance of one isolated intersection is determined by its own signal timings but also nearby signals. Research on signal coordination problems can be found in numerous literatures. A series of models, based on MAXBAND model, have been proved to be one effective approach for this purpose. Morgan and Little [9] first proposed the concept of bandwidth. A mixed-integer linear program (MILP) was formulated maximization model maximizes two-way green bandwidths of a given arterial so that vehicles may have larger chances to traverse the arterial without any stops. Later, the model was completed by Little et al. [10] with the consideration of lead or lag pattern for left turn phases and a queue clearance time, which is more general and called MAXBAND model. Several extensions were later proposed by other researchers, which enriched this model system. Chang et al. [11] extended MAXBAND model, and the new one was called MAXBAND-86, which can be used for solving multiarterial closed network problems. Gartner et al. [12] proposed MULTIBAND model, which can generate variable bandwidth progression schemes in which each directional road section is assigned an individually weighted band. Stamatiadis and Gartner [13] then extended the model to MULTIBAND-96, which was applied to solve arterial networks problems.

MAXBAND and MULTIBAND models offer traffic engineers a standard procedure of signal timing. Meanwhile, further research was conducted by interested researchers for extended consideration and specific scenarios. Lin et al. [14] introduced the maximum number of intersections that vehicles can traverse through in a signal progression band, the principle of the arrival sequences and the variations in the leading and trailing of adjacent bands to the model to resolve different traffic situations for arterial intersections. Zhang et al. [15] proposed an Asymmetrical MULTIBAND (AM-BAND) model is developed by relaxing the symmetrical progression band requirement in MULTIBAND. Yang et al. [16] presented a multi-path progression model for multiple critical path-flows contributing to the high volume in each arterial link. Zhang et al. [17] proposed two models, denoted as MaxBandLA and MaxBandGN, to tackle traffic signal coordination problems for long arterials and grid networks based on Little's bandwidth maximization model. Cho et al. [18] developed two path-based MAXBAND models by relaxing the assumptions of lengths of

the common cycle, green splits, and the free-flow speed. Multiple mode was considered by Ma et al. [19]. A partition-enabled multi-mode band (PM-BAND) model is designed to solve the signal coordination problem for arterials with passenger cars and transit vehicles.

From above, it can be seen that most models were designed for coordinating the through movement on the arterials. By applying these strategies, benefit of the minor arterial that intersects arterial is sacrificed. With the increasing of vehicle trips in the urban area, the proportion of flow on the minor arterial might be increased too. In this case, it might lead to decline of traffic efficiency by only considering through movement on the arterials. To figure out this problem, a numerical test is conducted in this paper by considering different proportion of flow on the minor arterial. MAXBAND method, a classical signal coordination model, is used for determining signal timing plans of different intersections on the arterial. The remainder of the paper is organized as follows: Sect. 2 introduces the formulation process of MAXBAND model. Numerical tests are presented in Sect. 3. At last, Sect. 4 is the conclusion.

2 Introduction to MAXBAND Formulation

MAXBAND model, developed by Little et al., is a basic and general method in arterial signal coordination. Though the extended model based on MAXBAND has investigated various coordination scenarios, the proportion of flow on the minor arterial was not considered. Variation tendency of traffic efficiency with different proportion might be similar between these models. Thus, MAXBAND model, the most classic model for signal coordination, is covered here as an analytical tool later. Given an arterial with fixed number of signals, MAXBAND model optimizes cycle time, offsets, speeds, and order of left turn phases for these signals to maximize the weighted combination of bandwidths.

Figure 1 shows green bands for both inbound and outbound direction between signals S_h and S_i. For signals S_h and S_i, there are two adjacent signals. The following Table 1 gives notations that will be used in formulation.

The formulation process is listed as Eqs. (1)–(12) in the following. The objective function Eq. (1) is to maximize the sum of the two-way green bandwidths, where $b, \bar{b}, w, \bar{w}, t, \bar{t}$ and τ represent vectors whose elements are the scalar decision variables $b_i, \bar{b}_i, w_i, \bar{w}_i, t_i, \bar{t}_i$ and τ_i. Equation (2) determines the progression preference to either the inbound or outbound direction. Equation (3) shows the lower and upper limits on cycle length. Equations (4) and (5) are constraints on green bandwidths to ensure the green bandwidth to be within the available green time. The constraint (6) is the loop integer constraint, which is designed to guarantee that the signals will not cause traffic flows to stop in the green bands. Equations (7)–(10) limit the variation of travel times.

$$\text{Max } b + k\bar{b} \tag{1}$$

Distance
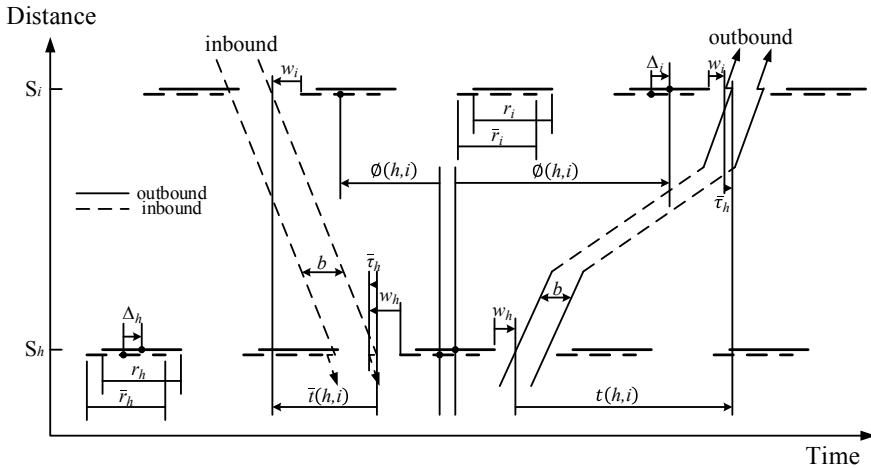
Fig. 1 Space–time diagram showing green bands for both inbound and outbound direction between signals S_h and S_i

Table 1 Notations in the formulation

Notations	Explanation
$b/\overline{b}$	outbound (inbound) bandwidth (cycles)
S_i	ith signal, $i = 1, \ldots, n$
$r_i/\overline{r}_i$	outbound (inbound) red time at S_i (cycles)
$w_i/\overline{w}_i$	time from right (left) side of red at S_i to left (right) edge of outbound (inbound) green band (cycles)
$t(h,i)/\overline{t}(h,i)$	travel time from S_i to S_h outbound [S_h to S_i inbound] (cycles)
$\Phi(h,i)/\overline{\Phi}(h,i)$	time from center of an outbound (inbound) red at S_h to the center of a particular outbound (inbound) red at S_i, the two reds are chosen so that each is immediately to the left (right) of the same outbound (inbound) green band
Δ_i	time from center of $\overline{r}_i$ to the nearest center of r_i. Positive if center of r_i is to right of center of $\overline{r}_i$ (cycles)
$\tau_i/\overline{\tau}_i$	queue clearance time, an advance of the outbound (inbound) bandwidth upon (leaving) S_i (cycles)
T	Cycle length (signal period) (seconds)
z	$(1/T)$ signal frequency (cycles/second)
T_1, T_2	lower and upper limits on cycle length
$d(h,i)/\overline{d}(h,i)$	distance between S_h and S_i outbound (inbound) (meters), i.e., $d_i = d(i, i+1), \overline{d}_i = \overline{d}(i, i+1)$
$e_i, f_i(\overline{e}_i, \overline{f}_i)$	lower and upper limits on outbound (inbound) speed (meters/second)
$1/h_i, 1/g_i$ $(1/\overline{h}_i, 1/\overline{g}_i)$	lower and upper limits on change in outbound (inbound) reciprocal speed, i.e., $1/h_i \leq 1/v_{i+1} - 1/v_i \leq 1/g_i$

$$\text{st. } (1 - k)\overline{b} \geq (1 - k)kb \tag{2}$$

$$1/T_2 \leq z \leq 1/T_1 \tag{3}$$

$$w_i + b \leq 1 - r_i \quad \forall i = 1, \ldots, n \tag{4}$$

$$\overline{w}_i + \overline{b} \leq 1 - \overline{r}_i \quad \forall i = 1, \ldots, n \tag{5}$$

$$(w_i + \overline{w}_i) - (w_{i+1} + \overline{w}_{i+1}) + (t_i + \overline{t}_i) + \delta_i l_i - \overline{\delta}_i \overline{l}_i - \delta_{i+1} l_{i+1} + \overline{\delta}_{i+1} \overline{l}_{i+1} - m_i$$
$$= (r_{i+1} - r_i) + (\tau_i + \tau_{i+1}) \quad \forall i = 1, \ldots, n - 1 \tag{6}$$

$$(d_i/f_i)z \leq t_i \leq (d_i/e_i)z \quad \forall i = 1, \ldots, n - 1 \tag{7}$$

$$(\overline{d}_i/\overline{f}_i)z \leq \overline{t}_i \leq (\overline{d}_i/\overline{e}_i)z \quad \forall i = 1, \ldots, n - 1 \tag{8}$$

$$(\overline{d}_i/h_i)z \leq (\overline{d}_i/\overline{d}_{i+1})t_{i+1} - t_i \leq (\overline{d}_i/g_i)z \quad \forall i = 1, \ldots, n - 2 \tag{9}$$

$$(\overline{d}_i/\overline{h}_i)z \leq (\overline{d}_i/\overline{d}_{i+1})\overline{t}_{i+1} - \overline{t}_i \leq (\overline{d}_i/\overline{g}_i)z \quad \forall i = 1, \ldots, n - 2 \tag{10}$$

$$b, \overline{b}, z, w_i, \overline{w}_i, t_i, \overline{t}_i; \geq 0 \tag{11}$$

$$m_i \text{integer}; \delta_i, \overline{\delta}_i \text{ binaryintegers} \quad \forall i = 1, \ldots, n \tag{12}$$

3 Numerical Tests

3.1 Simulation Scenario

A simulation scenario with an east–west arterial consists of seven signal intersections is created for simulation. The free-flow speed of the whole arterial was set to be 50 km/h. The geometric information and the original demand scenario are extracted in Zhang et al.'s paper [10], which is listed in Table 2. And the green splits of different phases and signal timing plans will be calculated based on Webster's theory. This initial state will be treated as the reference group to test arterial's traffic efficiency with increasing flow proportion of the minor arterials intersecting arterials. In the following subsections, the effect of increasing minor arterial flow on arterial coordination is teated, in which both left turn flow and through movement flow from

Table 2 Geometric information and original demand of tested arterial

Signal number	Distance (m)	East bound			West bound			South bound			North bound		
		LT	TH	RT	LT	TH	RT	LT	TH	RT	LT	TH	RT
1	–	62	1409	82	56	1361	138	40	200	52	38	219	109
2	341	85	1254	42	131	1291	114	24	179	42	54	150	51
3	695	50	1084	85	98	908	68	15	150	38	31	168	75
4	594	50	1328	68	90	1426	128	50	500	48	40	480	50
5	533	133	1245	69	90	1342	48	50	180	50	50	180	50
6	296	32	1053	100	85	1246	120	50	800	120	42	650	62
7	497	59	1095	156	100	1200	96	75	350	123	52	520	48

Note Unit of flow is pcu/h

minor arterial are considered. Synchro 7 is used for calculation and simulation. All the MILP problems are solved in Lingo 18.0 which is operated on a PC with 2.7 GHz CPU, 8 GB RAM, and Windows 10 operating system.

3.2   The Effect of Increasing Flow on the Minor Arterials

The effect of increasing through movement flow:

In this test, two intersections, numbered 1 and 5, are selected to adjust the flow for the four directions. To keep the total flow of tested intersections constant, the through movement flow on the minor arterials is creased. Four scenarios are designed with through movement flow increased 80 pch/h each time for test. The detailed demand information is listed in Table 3. And the results are shown in Table 4. The average delay for vehicles on the arterials, minor arterials, and entire network are presented, as well as two-way bandwidth.

The effect of increasing left turn flow:

In this subsection, the effect of increasing left turn flow is simulated. Similar with the through movement analysis, four scenarios (5–8) are also designed as listed in Table 5. And the simulated results are presented in Table 6.

Summary

From the simulated results, it can be easily observed that coordination efficiency is affected by the proportion of different direction flow on the arterials significantly. For the effect of through movement flow on the minor arterials, it can be seen that average delay for vehicles on the arterials increases with the increasing of proportion of through movement flow, as well as the entire average delay. Average delay for vehicles on the minor arterials and two-way bandwidth decreases at the same time. The results for effect of left turn flow are similar. It can be concluded that the increasing of flow on the minor arterials negative effect on the arterial coordination efficiency.

4   Conclusions

In this paper, the impact of increasing minor arterial flow for arterial coordination efficiency is analyzed. MAXBAND model is applied for optimizing signal plans of the tested arterials, in which two-way bandwidth can be obtained separately. An numerical test is conducted by designing eight scenarios with different proportion of through movement and left turn flow on the minor arterials. The results show that when the proportion of minor arterial flow increases, the arterial coordination efficiency would be reduced, especially for increasing left turn flow. This work reveals the improvement direction for arterial coordination. The effect of minor arterial flow will be considered in the further study about arterial coordination.

Table 3 Detailed information of tested scenarios 1–4

Scenario number	Signal number	East bound			West bound			South bound			North bound		
		LT	TH	RT	LT	TH	RT	LT	TH	RT	LT	TH	RT
1	1	62	1329	82	56	1281	138	40	280	52	38	299	109
	5	133	1165	69	90	1262	48	50	260	50	50	260	50
2	1	62	1249	82	56	1201	138	40	360	52	38	379	109
	5	133	1085	69	90	1182	48	50	340	50	50	340	50
3	1	62	1169	82	56	1121	138	40	440	52	38	459	109
	5	133	1005	69	90	1102	48	50	420	50	50	420	50
4	1	62	1089	82	56	1041	138	40	520	52	38	539	109
	5	133	925	69	90	1082	48	50	500	50	50	500	50

Note Unit of flow is pcu/h

Table 4 Results for scenarios 1–4

Scenario number		Delay/vehicle (s)	Two-way bandwidth
Reference	Entire network	26.42	28 + 34
	Arterial	15.29	
	Minor arterial	61.39	
1	Entire network	26.3	29 + 34
	Arterial	15.68	
	Minor arterial	57.56	
2	Entire network	27.77	29 + 31
	Arterial	16.98	
	Minor arterial	57.15	
3	Entire network	30.06	24 + 33
	Arterial	21.89	
	Minor arterial	51.26	
4	Entire network	26.92	27 + 28
	Arterial	17.71	
	Minor arterial	49.42	

Table 5 Detailed information of tested scenarios 5–8

Scenario number	Signal number	East bound			West bound			South bound			North bound		
		LT	TH	RT	LT	TH	RT	LT	TH	RT	LT	TH	RT
5	1	62	1329	82	56	1281	138	120	200	52	118	219	109
	5	133	1165	69	90	1262	48	130	180	50	130	180	50
6	1	62	1249	82	56	1201	138	200	200	52	198	219	109
	5	133	1085	69	90	1182	48	210	180	50	210	180	50
7	1	62	1169	82	56	1121	138	280	200	52	278	219	109
	5	133	1005	69	90	1102	48	290	180	50	290	180	50
8	1	62	1089	82	56	1041	138	360	200	52	358	219	109
	5	133	925	69	90	1022	48	370	180	50	370	180	50

Table 6 Results for the four scenarios

Scenario number		Delay/vehicles (s)	Two-way bandwidth
Reference	Entire network	26.42	28 + 34
	Arterial	15.29	
	Minor arterial	61.39	
5	Entire network	26.49	31 + 29
	Arterial	18.22	
	Minor arterial	50.82	
6	Entire network	27.78	24 + 33
	Arterial	18.96	
	Minor arterial	52.13	
7	Entire network	27.3	30 + 25
	Arterial	18.99	
	Minor arterial	48.86	
8	Entire network	31.72	25 + 27
	Arterial	19.49	
	Minor arterial	61.67	

References

1. Weng, J., Xue, S., Yang, Y., Yan, X., Qu, X.: In-depth analysis of drivers' merging behaviour and rear-end crash risks in work zone merging areas. Accid. Anal. Prev. **77**, 51–61 (2015)
2. Xu, C., Yang, Y., Jin, S., Qu, Z., Hou, L.: Potential risk and its influencing factors for separated bicycle paths. Accid. Anal. Prev. **87**, 59–67 (2016)
3. Qu, X., Yu, Y., Zhou, M., Lin, C.T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach. Appl. Energy **257**, 114030 (2020)
4. Zhou, M., Yu, Y., Qu, X.: Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: A reinforcement learning approach. IEEE Trans. Intell. Transp. Syst. **21**(1), 433–443 (2020)
5. Easa, S.M., Qu, X., Dabbour, E.: Improved pedestrian sight distance needs at railroad-highway grade crossings. J. Transp. Eng. Part A: Syst. **143**(7), 04017027 (2017)
6. Bie, Y., Cheng, S., Easa, S., Qu, X.: Stop line set back at a signalized roundabout: A new concept for traffic operations. J. Transp. Eng. ASCE **142**(3), 05016001 (2016)
7. Jin, S., Qu, X., Zhou, D., Xu, C., Ma, D., Wang, D.: Estimating cycleway capacity and bicycle equivalent unit for electric bicycles. Transp. Res. Part A: Policy Pract. **77**, 225–248 (2015)
8. Xu, J.: Traffic management and control, 1st edn. China Communications Press, Beijing (2007)
9. Morgan, J., Little, J.: Synchronizing traffic signals for maximal bandwidth. Oper. Res. **12**(6), 896–912 (1964)
10. Little, J., Kelson, M., Gartner, N.: MAXBAND: A program for setting signals on arteries and triangular networks. Transp. Res. Rec. **795**, 40–46 (1981)
11. Chang, E., Cohen, S., Liu, C., Chaudhary, N., Messer, C.: MAXBAND-86: Program for optimizing left-turn phase sequence in multiarterial closed networks. Transp. Res. Rec. **1181**, 61–67 (1988)
12. Gartner, N., Assmann, S., Lasaga, F., Hous, D.: MULTIBAND-a variable-bandwidth arterial progression scheme. Transp. Res. Rec. **1287**, 212–222 (1990)

13. Stamatiadis, C., Gartner, N.: MULTIBAND-96: a program for variable-bandwidth progression optimization of multiarterial traffic networks. Transp. Res. Rec. **1554**(1), 9–17 (1996)
14. Lin, L., Tung, L., Ku, H.: Synchronized signal control model for maximizing progression along an arterial. J. Transp. Eng. **136**(8), 727–735 (2009)
15. Zhang, C., Xie, Y., Gartner, N., Stamatiadis, C., Arsava, T.: AM-band: an asymmetrical multi-band model for arterial traffic signal coordination. Transpo. Res. Part C: Emerg. Technol **58**, 515–531 (2015)
16. Yang, X., Cheng, Y., Chang, G.: A multi-path progression model for synchronization of arterial traffic signals. Transp. Res. part C: Emerg. Technol **53**, 93–111 (2015)
17. Zhang, L., Song, Z., Tang, X., Wang, D.: Signal coordination models for long arterials and grid networks. Transp. Res. Part C: Emerg. Technol. **71**, 215–230 (2016)
18. Cho, H., Huang, T., Huang, C.: Path-based MAXBAND with green-split variables and traffic dispersion. Transp. B: Transp. Dyn. **7**(1), 726–740 (2019)
19. Ma, W., Zou, L., An, K., Gartner, N., Wang, M.: A partition-enabled multi-mode band approach to arterial traffic signal optimization. IEEE Trans. Intell. Transp. Syst. **20**(1), 313–322 (2018)

Traffic Safety Assessment of Deceleration Function Area Based on TTC Model

Weiwei Qi, Zhexuan Wang, and Bin Shen

Abstract Sections of tunnel entrances, industrial and mining schools with deceleration function zones are high-traffic zones due to their special traffic conditions. The instability of the car during the deceleration process and the driver's wrong deceleration operation may be important causes of traffic accidents. In order to improve the driving safety in the road deceleration function zone, the traffic flow at the entrance to Tianhe North Tunnel in Guangzhou City is taken as the research object, and we evaluate the traffic safety in the road deceleration function zone. The results show that speed standard deviation is a good predictor of potential risks, and speed standard deviation can be used to actively assess road safety. The research results help to further to optimize the driving behavior in the deceleration functional area and improve the safety of traffic flow in the deceleration functional area.

Keywords Deceleration functional area · Potential collision risk · Traffic safety assessment

1 Introduction

For some roads with poor road conditions or complex environment, because they have special sections such as poor horizontal and vertical alignment, long downhill road and too many tunnels, some countries will set up road deceleration functional areas to ensure the safety of drivers. Road deceleration function area refers to the area where speed limit measures are implemented on some road sections with poor road conditions or complicated environments to ensure driving safety. Because the deceleration function zone is always set in a more dangerous zone, the driving safety of drivers passing through this zone is poor. This issue should be paid more attention by scholars at home and abroad [1]. For example, in 2018, a traffic accident occurred on a section of the West Han Expressway in China, which resulted in 36 deaths.

W. Qi (✉) · Z. Wang · B. Shen
School of Civil Engineering and Transportation, South China University of Technology, 510641 Guangzhou, China
e-mail: ctwwqi@scut.edu.cn

The root cause of this accident is that the driver of the bus was driving dangerously in the deceleration function area in front of the tunnel and hit the front wall of the tunnel entrance. Such traffic accidents are still common in our country, so it is the key problem that traffic experts should pay attention to ensure the safety of road deceleration function area.

There are many factors that affect driving safety, such as fatigue driving, speeding and drunk driving. [2]. For example, a study by Violanti J. M. showed that drivers using mobile phones while driving performance degradation is an important cause of traffic accidents [3]. In general, there are three main factors affecting driving safety: First, the driver's body function is unbalanced due to fatigue. Second, the driver is distracted due to interference from external factors, that is, distracted driving. The third is the temporary damage of the driver, such as the reduction of the driving ability of the driver due to the side effects of drugs and alcohol [4]. The decline of the driver's driving ability is firstly manifested in the fact that the driver is easy to overspeed and his vision are reduced [5]. Studies have shown that following the guidance of induced anger, drivers will pass more yellow traffic lights ($P < 0.01$) and tend to drive faster [6]. The second is that the vehicle cannot be steered correctly, or the response time for operating the vehicle is increased. Henry found that the in-car concert distracted the driver through a simulated driving test. Both low-volume and high-volume music increased the driver's operating error rate [7]. These phenomena mean that the driver is in a dangerous driving state and prone to traffic accidents.

The driving state of the vehicle for the deceleration function zone is a specific manifestation of the driver's handling of the vehicle, and it can be determined whether the driver is in a driving dangerous state [8]. An important feature of the driver in a dangerous driving state is the increased response time of the vehicle [9]. Secondly, when the driver in dangerous driving state meets the speed limit sign, he often stops and decelerates in an emergency. Therefore, excessive acceleration is a sign of poor driving conditions. In addition, when the vehicle is passing through the deceleration function area, if the speed is too large and the speed is not reduced below the prescribed speed, it constitutes overspeed driving, which is also a dangerous driving state [10].

In order to study the driving safety of drivers during driving, many scholars at home and abroad have designed different traffic safety assessment models. Meng, Q. et al. developed a method for calculating vehicle collision frequency in highway tunnels by using negative binomial regression model [11]. Xu, C. et al. set up a generalized linear model and concluded that the risk increases with the increase of bicycle lane width and proportion of e-bikes. [12]. Kuang, Y et al. propose an alternative measure called Aggregated Impact Index (ACI) to measure the impact risk. The results show that the ACI is superior to the traditional three alternatives (impact time, parking distance ratio and impact risk index) in describing the risk of rear end collision [13]. Derbel O. evaluates the safety of time and distance standards by extending the dwell time. [14]. These models suggest ways to improve driver safety from different perspectives.

In addition, many scholars have turned their attentions to the future and started to study the traffic problems in terms of Internet of vehicles and pilotless driving.

Xiaobo Qu et al. developed a follow-up model for electric, connected and autonomous vehicles based on enhanced learning, designed to reduce traffic fluctuations caused by drivers and reduce power consumption [15]. The results of Manit and Surachate show that the presence of autonomous vehicles can improve the overall speed and traffic flow of the system and delay traffic congestion [16]. Mofan Zhou et al. developed a collaborative intelligent driver model and found that an increase in the percentage of autonomous vehicles will reduce total driving time and mitigate traffic fluctuations [17]. Sangmin L. et al. have developed a traffic control system based on machine learning prediction to solve the actual traffic congestion problem of large factories [18]. Mofan Zhou et al. proposed a microcar following model based on recurrent neural network, which can accurately capture and predict traffic oscillations for predicting traffic fluctuations in different driver characteristics [19]. These studies have a very enlightening significance of the application prospects of traffic safety assessment models.

In summary, the driving safety issues in the deceleration function area are quite complicated. The main factors affecting driving safety are fatigue driving, distracted driving and drunk driving. And the zone where the deceleration function zone set is often a high-incidence zone for road traffic accidents. Therefore, when the driver passes the road deceleration function zone, dangerous driving behaviors such as excessively long reaction time, sudden braking or speeding are prone to occur, thereby greatly reduces driving safety. Many scholars have proposed traffic safety assessment models from different angles in order to find a good way to improve driving safety. In addition, some scholars' research papers on the Internet of vehicles and driverless vehicles provide broader ideas for the study of traffic safety.

## 2   Traffic Safety Assessment Theory

Two or more road users always have a certain distance in a certain time and space. At this time, if their operating status is not changed, there is a risk of collision, which is a traffic conflict [20]. Traffic conflicts are usually reflected on time-to-collision (TTC) values. TTC refers to the remaining time before collision of two vehicles if the collision process and speed difference remain the same on the road sections, which can be expressed in Fig. 1 and the following formula [21].

$$\text{TTC} = \begin{cases} \frac{L_{\text{leader}} - L_{\text{follower}} - L_{\text{leader}}}{\dot{L}_{\text{follower}} - \dot{L}_{\text{leader}}}, & \text{if } \dot{L}_{\text{follower}} > \dot{L}_{\text{leader}} \\ \infty & \text{Otherwise} \end{cases} \tag{1}$$

Among them, l_{leader} is the length of the leader vehicle in a specific area, L_{leader} and L_{follower} are the positions of the leader and follower vehicles in a specific area and $\dot{L}_{\text{leader}}$ and $\dot{L}_{\text{follower}}$ are the speeds of the leader and follower vehicles in a specific area [19]. The TTC measurement procedure for target vehicle following is as follows:

(1)   The specific area in this experiment refers to the deceleration function area.
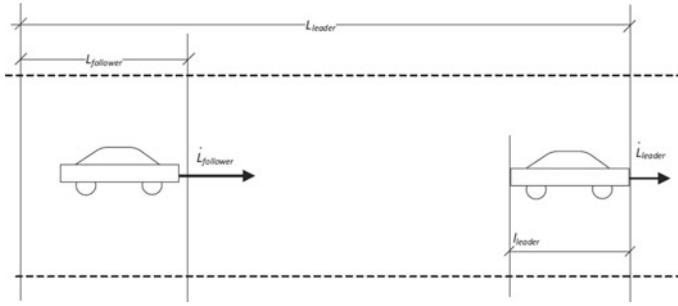
Fig. 1 Vehicle potential collision model

(2) In the experimental video, the length of the two vehicles in front of the specific area is measured by the image recognition equipment: l_{leader}, the point speed of the two vehicles: $\dot{L}_{leader}$ and $\dot{L}_{follower}$, and the position of the two vehicles at this time L_{leader} and $L_{follower}$.

(3) TTC can be calculated according to Formula (1) for the specific vehicle following. Note that this estimate is based on the assumption that the average velocity can be approximated by the point velocity.

If TTC value is less than a certain threshold, it can be regarded as a road traffic conflict. Regarding the threshold, this article refers to some opinions on the literature. For example, Hirst and Graham (1997) used 4 s as TTC threshold in their reports to distinguish the situation in which the driver inadvertently found himself in danger state and the situation in which the driver was still in control state [22]. The minimum value of unsupported driver proposed by Hogema and Janssen (1996) was 3.5 s, and the minimum value of supported driver was 2.6 s [11, 23]. Because the amount of data obtained in this paper is not sufficient, in order to make the risk index more obvious and easier to analyze, 4 s is used as the TTC threshold.

3 Experiment Design and Data Collection

3.1 Demonstration of Actual Test in Deceleration Function Area

This experiment adopts the video observation method to record the deceleration function area of Guangzhou Tianhe North Tunnel entrance and the traffic conditions of ordinary roads nearby, and provides data support for safety assessment and traffic to flow modeling. The experimental equipment mainly includes cameras, ice cream cones, roller rangefinders and other equipment. The roller rangefinder can measure the length to determine the placement of the detection point, and place the ice cream cone at the detection point as a mark. The camera is used to collect data such as

vehicle position, speed in different lanes of the experimental section and the control section.

3.2 Participant and Procedure

This experiment requires two equipment operators and three observers. The equipment operators are mainly responsible for operating the camera to shoot the traffic conditions on the experimental section. The observers are responsible for inspecting whether the conditions of the speed limit section meet the experimental requirements, and selecting appropriate observation points and shooting points.

The test site was selected as the entrance section of the Tianhe North Tunnel, and the entire line (including the speed limit front section) was a two-way six-lane road to a speed limit of 40 km/h in front of the tunnel. The straight section of the Tianshou Road where the tunnel is located can accelerate the road section for a long time, so the vehicle speed is relatively fast. In addition, the road in front of the tunnel entrance is a longitudinal slope, which may cause a potential collision risk. There are fewer obstacles on the road section, and there is a pedestrian bridge. So the observation points and detection points are easy to choose and it is convenient to shoot (Figs. 2 and 3).

After the start of the experiment, the experimenter starts to take field photos of the entrance and middle road section for more than one hour, respectively, and obtains the average speed data, vehicle position and other data between each two detection points through computer processing. Some initial data are shown in Table 1.

Fig. 2 Real picture of section before entrance of deceleration function area

Fig. 3 Real picture of entrance section of deceleration function area

Table 1 Some vehicle data

Vehicle group	Vehicle number	Lane (1/2/3)	Model (Big/Small)	Speed 1 (m/s)	Speed 2 (m/s)	Vehicle spacing (m)
1	1	1	Small	9.09	9.09	8.16
2	2	1	Small	10.00	8.70	4.24
3	3	1	Small	9.09	8.70	11.00
4	4	1	Small	10.53	9.52	16.00
5	5	1	Small	9.52	9.52	8.57
6	6	1	Small	11.11	10.53	20.73
7	7	1	Small	9.09	11.11	15.19
…	…	…	…	…	…	…

4　Traffic Safety Evaluation Model

In the experiment, we first explore the impact of different lanes and different road sections on traffic safety, and then take this as a starting point to analyze the influencing factors of driving safety, and find a model that can evaluate road traffic safety. We propose the potential collision risk from the perspective of the transportation department, and evaluate the road safety according to the frequency of collision and conflict in a specific section of road in a specific period of time. From this perspective, risk is interpreted as the ratio of the number of conflicts to time, which is mathematically [21].

$$\text{SR}_i = \frac{N_{i,\text{TTC}<\tau}}{t} \tag{2}$$

Table 2 Potential collision risks

Sample	Road	Lane	Risk
1	1	1	0.06
2	1	2	0.04
3	1	3	0.12
4	2	1	0.16
5	2	2	0.2
6	2	3	0.26

In the Formula (2), $N_{i,\text{TTC}<\tau}$ indicates that the number of TTC samples is smaller than that in the section i, and t is the time period of the survey. This paper refers to this risk as social risk. According to the above, the value is 4 s.

The observation sections are divided into two categories: the internal sections of the deceleration function area and the entrance sections of the deceleration function area. For each road section, the risks of different lanes (outside lane, middle lane and inside lane) are calculated according to Formula (2), and the results are shown in Table 2.

Mark:
Section type: 1-entrance to functional area; 2-internal functional area,
Lane types: 1-inside lane; 2-middle lane; and 3-outside lane.

We used single-factor analysis of variance (ANOVA) combined with post-hoc tests to analyze the impact on road type on risk. First, we test the homogeneity of variance. In the 'one-factor homogeneity test' we get $P = 0.812 > 0.05$, which shows that the variance is homogeneous, and one-way analysis of variance can be performed. Then, SPSS software was used to perform the one-factor ANOVA on the data, and $P = 0.024 < 0.05$ was obtained, indicating that there is a significant difference between the two groups of data. The specific results are shown in Table 3.

Based on the results, we draw the following conclusions: the potential collision risk at the entrance of the road deceleration functional area is greater than that inside the deceleration functional area (P value $= 0.024 < 0.05$). The average potential collision risk is 0.07 and 0.21, which means that the internal section is relatively safer.

One-factor analysis of variance and post-hoc tests can also be used to test the risk impact on lane types. In the risk impact analysis of the deceleration function area, the dependent variable is risk, and the only variable factor is the type of lane. SPSS software was used to perform an one-way analysis of variance in the data, and $P =$

Table 3 Variance analysis results of potential collision risks on different road sections

Road	P value	Risk average
1	0.024 < 0.05	0.07
2		0.21

Table 4 Analysis of variance of potential collision risks in different lanes

Road	P value	Risk average
1&2	0.924 > 0.05	0.11&0.12
1&3	0.465 > 0.05	0.11&0.19
2&3	0.518 > 0.05	0.12&0.19

$0.694 > 0.05$ was obtained its relationship of potential collision risk cannot be found from the perspective of lane types. The results are shown in Table 4.

Considering that the lanes in the deceleration function area are all immutable lanes, the vehicle will not be affected by vehicle lane change during the operation. Therefore, we preliminarily judge that the difference in lane types has no significant relationship to the potential collision risk of the vehicle.

At this time, we realize that in different road sections, there may be a factor that affects driving safety, resulting in a higher potential collision risk at the entrance of the deceleration function area than that in the deceleration function area. There are three possible indicators: speed, speed standard deviation and traffic volume. Table 5 shows the specific values of these three safety indicators. The scatter diagram of the three security indicators is shown in Fig. 4. However, there is no same rule between average speed and standard deviation of speed with traffic volume. In order to check which indicator has better performance, we will conduct an active safety assessment for driving in the deceleration function area.

The relationship between the potential collision risk and the three safety indicators was studied using a univariate linear regression analysis method, as shown in Table 6. Compared with speed and traffic volume, the standard deviation of speed provides the best prediction of potential car accident risk. In the risk-speed standard deviation model, the P value $= 0.003 < 0.05$, indicating that reducing the speed standard deviation can significantly reduce the potential collision risk. The $R2$ (determination coefficient) value of the model reached 0.91, indicating that the model fits well.

Figure 5 shows the relationship between the potential collision risk and the best predictor (standard deviation of speed). As shown, the standard deviation of speed usually has a linear trend of potential collision risk. This result indicates that the speed standard deviation is a better predictor of potential risks, and the speed standard deviation can be used to actively evaluate road safety. This conclusion also shows

Table 5 Macro-security index data

| Sample	Average speed	Standard deviation of speed	Traffic
1	10.60	1.21	1008
2	10.37	1.25	768
3	9.20	1.40	528
4	10.03	1.39	1044
5	10.11	1.70	924
6	10.10	1.83	540

Fig. 4 Scatter diagram of macro-security indicators
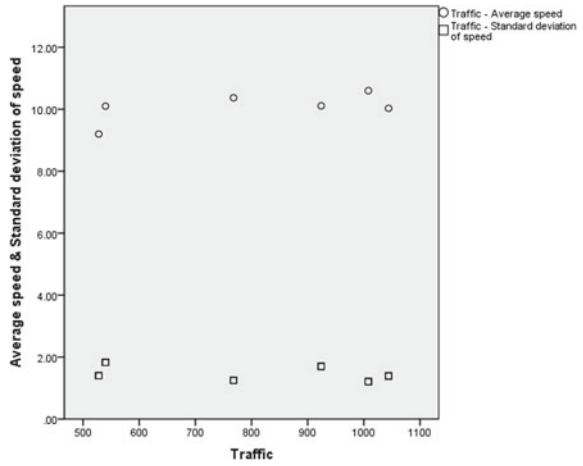
Table 6 Forecast results of micro safety indicators

Linear regression model	P value	$R2$
Risk—average speed	0.63 > 0.05	0.062
Risk—speed standard deviation	0.003 < 0.05	0.910
Risk—traffic	0.59 > 0.05	0.079

Fig. 5 Risk-speed standard deviation linear regression curve fitting
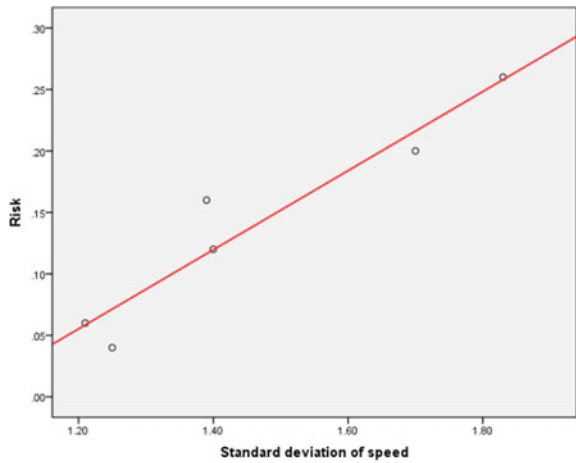
that the significant difference of potential collision risk between the interior and the entrance of deceleration function area is related to the difference of standard deviation of vehicle speed.

5 Summary

According to the above research, the potential collision risk of the entrance to the deceleration functional area is lower than that inside the deceleration functional area, and different lanes have no significant impact on the potential collision risk. Secondly, the research shows that the speed standard deviation is a better predictor of potential risks. This result explains why the entrance to the deceleration function area is more risky than the interior, and also proposes a more active safety assessment method than the potential collision risk, that is, using standard deviation of speed to evaluate road safety. This conclusion can be used as a safety improvement measure in the deceleration function zone under the mature conditions of the vehicle network and driverless technology, and it will be useful in the near future.

References

1. Yulong, P.: Road Traffic Safety, pp. 249. People's Communications Press, Beijing (2007)
2. Patel, M., Lala, S.K.L., Kavanagha, D., Rossiterb, P.: Applying neural network analysis on heart rate variability data to assess driver fatigue. Expert Syst. Appl. **38**(6), 7235–7242 (2011)
3. Violanti, J.M., Marshal, J.R.: Cellular phones and traffic accidents: an epidemiological approach. Accid. Anal. Prev. **28**(2), 265–270 (1996)
4. Yeo, M.V.M., Li, X.P., Shen, K., et al.: Can SVM be used for automatic EEG detection of drowsiness during car driving. Saf. Sci. **47**(1), 115–124 (2009)
5. Abdu, R., Shinar, D., Meiran, N.: Situational (state) anger and driving. Transp. Res. Part F: Traffic Psychol. Behav. **15**(5), 575–580 (2012)
6. Brodsky, H., Hakkert, A.S.: Risk of a road accident in rainy weather. Accid. Anal. Prev. **20**(3), 161–176 (1988)
7. Henry, E.L.: The effect of music volume on simulated interstate driving skills. The Florida State University College of Music, Florida (2006)
8. Sayed, R., Eskandarian, A.: Unobtrusive drowsiness detection by neural network learning of driver steering. Proc. Inst. Mech. Eng. Part D: J. Automob. Eng. **215**(9), 969–975 (2001)
9. Makoto, U., Akio, K., Hirotsugu, M., et al.: Fatigue analysis based on synthesis of psychological and physiological responses measured simultaneously in follow-up driving. J. East. Asia Soc. Transp. Stud. (EASTS) **6**, 3325–3340 (2005)
10. Chaovalit, P., Saiprasert, C., Pholprasit, T.: A method for driving event detection using SAX on smartphone sensors. In: Proceedings of 2013 13th International Conference on ITS Telecommunications, Tampere, pp. 450–455 (2013)
11. Meng, Q., Qu, X.: Estimation of vehicle crash frequencies in road tunnels. Accid. Anal. Prev. **48**, 254–263 (2012)
12. Xu, C., Yang, Y., Jin, S., Qu, Z., Hou, L.: Potential risk and its influencing factors for separated bicycle paths. Accid. Anal. Prev. **87**, 59–67. https://doi.org/10.1016/j.aap.2015.11.014
13. Kuang, Y., Qu, X., Wang, S.: A tree-structured crash surrogate measure for freeways. Accid. Anal. Prev. **77**, 137–148 (2015)

14. Derbel, O., Mourllion, B., Basset, M.: Extended safety descriptor measurements for relative safety assessment in mixed road traffic. In: 2012 15th International IEEE Conference on Intelligent Transportation Systems (ITSC) (2012)
15. Qu, X., Yu, Y., Zhou, M., Lin, C.T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach. Appl. Energy **257**, 114030 (2020)
16. Manit, K., Surachate, L.: Dissipation of traffic congestion using autonomous-based car-following model with modified optimal velocity. Phy. A: Statist. Mech. Appl. 123412 (2019)
17. Zhou, M., Qu, X., Jin, S.: On the impact of cooperative autonomous vehicles in improving freeway merging: A modified intelligent driver model based approach. IEEE Trans. Intell. Transp. Syst. **18**(6), 1422–1428 (2017)
18. Sangmin, L., Younghoon, K., Hyungu K., Soon-Kyo L., Seokhyun C., Taesu C., Keeyong S., Jeehyuk P., Seoung B.K.: Intelligent traffic control for autonomous vehicle systems based on machine learning. Expert Syst. Appl. **144** (2020)
19. Zhou, M., Qu, X., Li, X.: A recurrent neural network based microscopic car following model to predict traffic oscillation. Transp. Res. Part C **84**, 245–264 (2017)
20. Zheng, L., Sayed, T.: A full Bayes approach for traffic conflict-based before-after safety evaluation using extreme value theory. Accid. Anal. Prev. **131**, 308–315 (2019)
21. Qu, X., Yang, Y., Liu, Z., Jin, S., Weng, J.: Potential crash risks of expressway on-ramps and off-ramps: A case study in Beijing China. Saf. Sci. **70**, 58–62 (2014)
22. Hirst, S., Graham, R.: The format and presentation of collision warnings. Ergonomics Saf. Intell. Driver Interfaces **2**, 203–219 (1997)
23. Hogema, J.H., Janssen, W.H.: Effects of intelligent cruise control on driving behaviour : A simulator study. Soesterberg, The Netherlands. Report, TM-1996-C-12 (1996)

Realistic 5.9 GHz DSRC Vehicle-to-Vehicle Wireless Communication Protocols for Cooperative Collision Warning in Underground Mining

Abdellah Chehri, Hamou Chehri, Nadir Hakim, and Rachid Saadane

Abstract Industrial vehicle automation is a core component of the building Industry 4.0. The uses of self-driving vehicles, inspection robots, and vehicular ad hoc networks (VANETs) communications in the mining industry are expected to open significant opportunities for collecting and exchanging data, localization, collision warning, and up-to-date traffic to enhance both the safety of workers and increase the productivity. In this paper, we present a review of the large-scale fading channel at 5.9 GHz in confined areas. Then, the requirements for DSRC receiver performance for VANET applications in an underground mine is calculated. This paper also reports the overall performance evaluation of three existing routing protocols, namely, emergency message dissemination for vehicular environments (EMDV), enhanced multi-hop vehicular broadcast (MHVB), and efficient directional broadcast (EDB) for active safety applications. Finally, a comparative study of these three routing protocols for cooperative collision warning in underground mining galleries was evaluated.

Keywords VANET networks · Cooperative collision warning · Channel modeling · Underground mine · Channel measurement

---

A. Chehri (✉)
Department of Applied Sciences, University of Québec in Chicoutimi, Saguenay ,
QC G7H 2B1, Canada
e-mail: achehri@uqac.ca

H. Chehri
Bell-Canada, 671 Rue de la Gauchetière Ouest, Montreal, QC H3B 2M8, Canada
e-mail: hamou.chehri@uqat.ca

N. Hakim
LRTCS Laboratory, University of Quebec in Abitibi-Témiscamingue (UQAT), Val-D'Or,
QC J9P 6W6, Canada
e-mail: nadir.hakem@uqat.ca

R. Saadane
Laboratory Engineering System, SIRC/LAGeS-EHTP, El Jadida, Morocco
e-mail: saadane@ehtp.ac.ma

133

1 Introduction

VANETs have been identified as a promising intelligent transportation system (ITS) technology to improve traffic safety and efficiency. Vehicular networks consist of vehicles and all the entities with which cars can establish communications. This can give the possibility of three modes of operation. The first mode is called "infrastructure mode," where the vehicle connects to a fixed station for the acquisition or transmission of information; this mode is called "vehicle-to-infrastructure" or V2I communications. The ad hoc mode in which vehicles collaborate in a decentralized way without relying on any infrastructure and form an ad hoc network (vehicle-to-vehicle or V2V). The third mode, hybrid mode, combines the first two modes. In the last couple of years, those modes (i.e., V2V and V2I) have attracted the interest of many researchers across the world [1–3].

V2V and V2I have been deployed in various applications, e.g., safety [4], navigation [5], cooperative driving [6], and advanced cruise control [7, 8]. Another area of research that could be highly promising is the development of high-speed and autonomous driving of underground mining vehicles.

However, machines and underground mines vehicles have become progressively more significant and more powerful. Improving worker's safety has become a concern. Therefore, there is an increased risk of serious injury caused by humans having to work in such confined spaces with heavy machinery [9–11].

Moreover, the industrial vehicle is a huge truck that operates in a harsh environment, with safety figuring alongside profitability at the top of the criteria list [12–15].

The modeling of radio propagation has not attracted as much interest as that of mobility. The signals transmitted by the vehicles propagate in the external environment and are degraded by several obstacles whose effect is difficult to predict.

Mobile vehicle networks are, in theory, capable of using a large number of transmission technologies, including those from the third generation (3G), fourth generation (4G-LTE), and recently fifth generation (5G). Besides, VANET can use IEEE 802.11x, WiMax, and DSRC/WAVE standards. The VANETs could also combine one or more of these technologies.

Wireless access for vehicular environment (WAVE) is a set of particular standards, which has been developed by the IEEE research group for VANETs. The DSRC/WAVE technology is already used for specific automotive applications, such as electronic toll payments, without stopping. We will present this technology in detail. Dedicated short-range communications, DSRC, were created specifically for mobile vehicle networks. They include the WAVE standards as well as the 802.11p standard, which is the Wi-Fi standard for mobile vehicle networks. WAVE combines the IEEE 1609x standards (1, 2, 3, and 4), which are the standards for layers 2, 3, 4, and 7 of the Open System Interconnection (OSI) model. IEEE 802.11p manages the physical layer (layer 1) of this same model.

Even though many DSRC vehicular communication measurement campaigns were conducted up to now, there are still omitted environments and frequency bands,

which require to be intensively investigated. The earliest V2V and V2I channel work used channel classifications similar to those used for cellular, i.e., rural, suburb, and urban in open areas [16]. As V2V applications were studied further, additional environments, e.g., parking, street intersections [17], and tunnels [18], which have not received so much attention so far.

The contribution of this paper is threefold. First, a review of large-scale fading at 5.9 GHz in confined areas are provided, which have not been thoroughly investigated in our previous works [19–21]. Then, the requirements for DSRC receiver performance is calculated. Finally, a comparative study of three routing protocols for cooperative collision warning in underground mining galleries was evaluated.

The paper is organized as follows: in the next section, a state-of-art of the VANET in tunnels and confined areas is given. We give a brief overview of radio propagation in tunnels and confined areas in Sect. 2. Section 3 evaluates the received signal strength (RSS) for the DSRC standard. The MAC and routing protocols for VANETs are given in Sect. 4. Results and discussion are provided in Sect. 5. The conclusions and future works are drawn in Sect. 6.

2 VANET in Tunnels and Confined Areas

A survey of V2V communication is presented in [16] and [22], where it is shown that the channel characteristics vary depending on the type of road (highway, rural, suburban, and urban), which can be explained by the different velocities and surroundings.

Tunnels are confined environments, where the radio propagation differs from other open environments. Under the intelligent transportation systems concept, the understanding of the V2I and V2V channel models is critical for a successful deployment.

The V2V and V2I signal propagation measurements in a parking garage have been performed in [23], in road tunnels [24–27], and on-bridge environments [28].
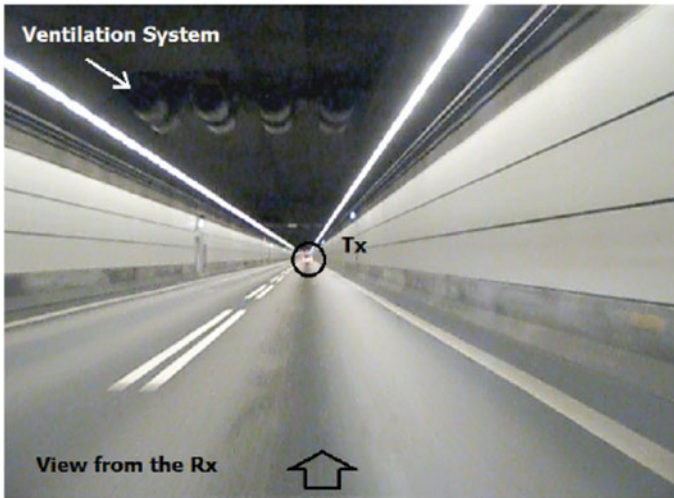
According to the literature, a tunnel and underground passageway could be considered a waveguide due to its geometry and conductivity [29]. However, the parameters that affect the radio propagation in tunnels differ from tunnels to other. This difference is related to the electromagnetic properties of the material used to build the tunnel, antenna characteristics, smoothness of the walls, radio obstacles, and particularly the tunnel's geometry [29].

Furthermore, the earlier research on radio propagation inside tunnels has shown that the dimensions of tunnels have a significant impact on the signal attenuation [29]. Moreover, the signal attenuation inside the tunnels depends also on radiation pattern, polarization, and position of transmuting and receiving antennas [30].

This work differs with these previous papers (i.e., [24–27]). The underground mines gallery is different from the road tunnels because the rough walls are higher than that in a tunnel with smooth walls (please see Fig. 1a, b).

Fig. 1 **a** Photography of the mine gallery (CANMET-Canadian Center for Minerals and Energy Technology). **b** Picture taken from Bernado et al. [25]

The environment in underground mines consists mainly of very rough walls, and the floor is not very flat. It is along these walls that cables and pipes are stretched near the ceiling.

The gallery where the measurements were done is located at a 70 m deep underground level. It stretches over a length of about 140 m with dimensions that vary between 2.5 and 3 m in width and a height of nearly 3 m.

Table 1 Simulation parameters

| Parameters	Value
Length of the mine tunnel	10 km
Frequency	5.9 GHz
Antenna type	Omni-directional
Channel model	Rayleigh fading channel
Packet size	512 bytes
Data rate	6 Mbps
Number of vehicles	20, 30, 50
Vehicle speed (km/h)	0–50 km/h
Communication distance vehicles	100 m
Warning message interval	1 s

3 Received Power Evaluation of DSRC

We utilize the statistical model described in the previous section to calculate the received signal attenuation. We use a large-scale model to estimate the received signal power for the DSRC application. We employed this model for its simplicity and the fact that it is well used in the literature.

For the analysis, we covered a range of 100 m. The transmit power set to 20 dBm, the antenna gain for both transmitters and receivers, was set to 3 dBi, at the 5.9 GHz frequency band.

The estimated received signal for two scenarios (vehicle moving of speed 1 m/s and the second vehicle is moving of speed on 5 m/s). By using the minimum sensitivity thresholds as defined in the DSRC standard. Based on the sensitivity thresholds presented in Table 1, we could determine the most suitable modulation and data rate for different vehicle speeds (as described in Fig. 2).

4 MAC and Routing Protocols for VANETs

At the MAC layer, it has been observed that the CSMA/CA mechanism used in many radio systems does not guarantee access to the channel before a specific random duration due to the back off time. Consequently, a significant delay can be generated, and the performance of the mechanism is degraded. However, in [31], for example, the authors propose solutions aimed at predicting access to the channel and therefore controlling this random duration. However, due to the wireless nature of vehicle networks, a road can suddenly abrupt, thus affecting the performance of the application. Consequently, the design of interlayer solutions can be beneficial in this type of network [32].
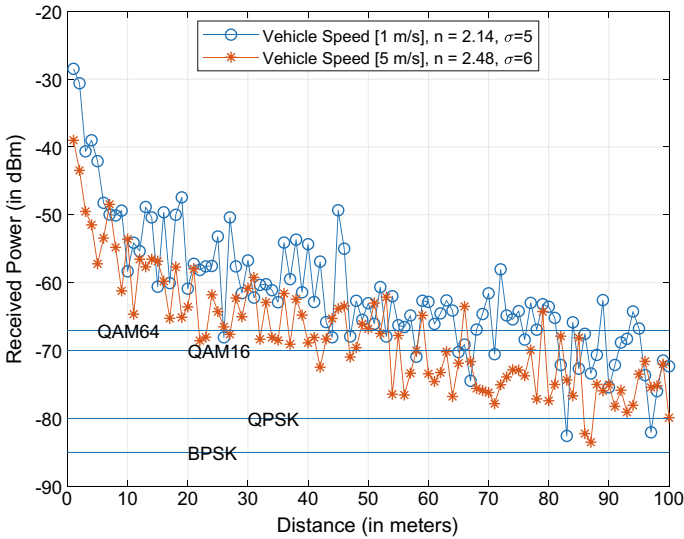
Fig. 2 Impact of vehicles speeds on the received signal power on underground mine galleries

The routing of information in mobile vehicle networks is not a simple problem. Indeed, VANETs are spontaneous networks, which must be able to set up independently.

The main challenge in the design of communication protocols in cooperative vehicle networks is to provide excellent network performance under the constraints of vehicle speed, intermittent connectivity, and rapid changes in topology.

These strong constraints make network organization mandatory if we want to provide advanced services to the user. Indeed, the organization into small groups of vehicles (clusters) makes it possible to optimize the data exchanges for the applications, which must function on this network.

These clusters are dynamic and are often involved in setting up virtual backbones. The organization of this network can be decentralized self-organization, in which the vehicles organize themselves to decide the topology of the clusters, or centralized, in which an entity external to the network of cars has the role of organizing clusters.

There are also a significant number of routing protocols, which have been proposed aiming to offer a high delivery rate, and reduced end-to-end communication time. In this work, we adopted three existing routing protocols, namely emergency message dissemination for vehicular environments (EMDV) [33], enhanced multi-hop vehicular broadcast (MHVB) for active safety applications [34], and efficient directional broadcast (EDB) [35]. EMDV uses the concept of a forwarding area, an area only in which stations are allowed to forward a safety message.

The MHVB applies an angle-based forwarder selection area referred to as a "backfire region" to select a potential forwarder. The EDB protocol uses the distance-based
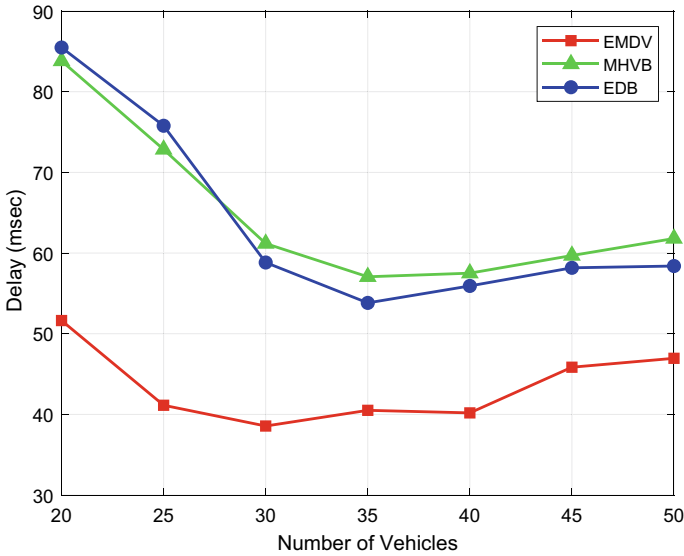
Fig. 3 Average delay as a function of the number of vehicles

forwarder selection to re-broadcast a received packet. When a vehicle receives a data packet, it waits for some time before re-broadcasting it [36].

5 Results and Discussion

In this section, we present the simulation VANET network in underground mines under various conditions for each of the performance parameters using the NetSim software tool and MATLAB environment for analysis and plotting.

The quadrature phase-shift keying (QPSK) modulation with 6 Mbps data rates is considered (as given in Table 1). During this initial interval, no messages are exchanged between vehicles. Additional simulator settings are presented in Table 1.

In Fig. 3, the average number of forwarders for all selected protocols increases by increasing the target region length, and it is almost the same for all the selected contracts.

6 Conclusion

In this work, simulation-based analysis has been carried out to analyze the VANET system performance using different routing protocols for cooperative collision warning in underground mining. This work is based on the NetSim software platform,

which takes several considerations of channel models, path losses, and wireless access technologies.

Our proposed model offers many key insights that can be used to enhance the overall performance of VANET systems for mining applications system. Results show that the performance of the VANET is improved by adopting EMDV routing protocol compared to MHVB, and EDB protocols in terms of average delay and number of potential forwarders.

References

1. Bhoi, S.K., Khilar, P.M.: Vehicular Communication—a survey. IET Netw. **3**(3), 204–217 (2014)
2. Alotaibi, M.M., Mouftah, H.T.: Relay selection for heterogeneous transmission powers in VANETs. IEEE Access **5**, 4870–4886 (2017)
3. MacHardy, Z., Khan, A., Obana, K., Iwashina, S.: V2X access technologies: regulation research and remaining challenges. IEEE Commun. Sur. Tut. **20**(3), 1858–1877 (2018)
4. Gokulakrishnan, P., Ganeshkumar. P: Road accident prevention with instant emergency warning message dissemination in vehicular ad-hoc network. PloS One **10**(5) (2015)
5. Yu, M.Y., Song, J., Zheng, K, Guo, Y: A beacon transmission power control algorithm based on wireless channel load forecasting in VANETs. PloS One **10**(11) (2015)
6. van Nunen, E., Kwakkernaat, R., Ploeg, J., Netten, B.D.: Cooperative competition for future mobility. Intell. Transp. Syst. IEEE Trans. **13**(3), 1018–1025 (2012)
7. Tang, T.-Q., et al.: An extended car-following model with consideration of the reliability of inter-vehicle communication. Measurement **58**, 286–293 (2014)
8. Kesting, M., Treiber, D., Helbing, D.: Connectivity statistics of store-and-forward intervehicle communication. Intell. Transp. Syst. IEEE Trans. **11**(1), 72–81 (2010)
9. Duff, E.S., Roberts, J.M., Corke, P.I: Automation of an underground mining vehicle using reactive navigation and opportunistic localization. In: Australasian Conference on Robotics and Automation, pp. 151–156. Auckland (2002)
10. Dragt, B.J: Modeling and control of an autonomous underground vehicle. University of Pretoria (2006)
11. Chehri, A., Fortier, P., Tardif, P.-M.: Security monitoring using wireless sensor networks. In: IEEE Communication Networks and Services Research, CNSR'07, pp. 13–17 (2007)
12. Chehri, A., Fortier, P., Tardif, P.M.: An investigation of UWB-based wireless networks in industrial automation. Int. J. Comput. Sci. Netw. Secur. **8**(2), 179–188 (2008)
13. El Ouahmani, T., Chehri, A., Hakem, N.: Bio-inspired routing protocol in VANET networks—a case study. In: Elsevier's Procedia Computer Science, 23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Budapest, Hungary (2019)
14. Karedal, J., Czink, N., Paier, A., Tufvesson, F., Molisch, A.F.: Path loss modeling for vehicle-to-vehicle communications. IEEE Trans. Veh. Technol. **60**(1), 323–328 (2011)
15. Cheng, L., Henty, B.E., Stancil, D.D., Bai, F., Mudalige, P.: Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 GHz dedicated short-range communication (DSRC) frequency band. IEEE J. Sel. Areas Commun. **25**(8), 1501–1516 (2007)
16. Molisch, A.F., Tufvesson, F., Karedal, J., Mecklenbrauker, C.F.: A survey on vehicle-to-vehicle propagation channels. IEEE Wirel. Commun. **16**(6), 12–22 (2009)
17. Schumacher, H., et al.: Vehicle-to-vehicle 802.11p performance measurements at urban intersections. In: Proceedings of IEEE ICC, Workshop on Intelligent Vehicular Networking, pp. 10–15. Ottawa, ON (2012)

18. Bernado, L., Roma, A., Paier, A., Zemen, T., et al.: In-tunnel vehicular radio channel characterization. In: Proceedings of IEEE Spring VTC, Budapest, Hungary, 15–18 May (2011)
19. Chehri, H., Chehri, A., Hakem, N.: In underground vehicular radio channel characterization. In: Elsevier's Procedia Computer Science, 23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Budapest, Hungary (2019)
20. Chehri, H., Chehri, A., Hakem, N.: Empirical radio channel characterization at 5.9 GHz for vehicle-to-infrastructure communication. In: IEEE 90th Vehicular Technology Conference, Hawaii, USA, 22–25 Sept (2019)
21. Chehri, H., Hakem, M.: Large scale propagation analysis of vehicle-to-vehicle communications at 5.9 GHz. In: IEEE Antennas and Propagation Society International Symposium (APSURSI), pp. 6–11 Memphis, Tennessee, USA (2014)
22. Viriyasitavat, W., Boban, M., Tsai, H.M., Vasilakos, A.: Vehicular communications: survey and challenges of channel and propagation models. IEEE Veh. Technol. Mag. **10**, 55–66 (2015)
23. Sun, R., Matolak, D.W., Liu, P.: 5 GHz V2V channel characteristics for parking garages. IEEE Trans. Veh. Technol. (2016)
24. Masson, E., Combeau, P., Berbineau, M., et al.: Radio wave propagation in arched cross section tunnels-simulations and measurements. J. Commun. **4**(4), 276–283 (2009)
25. Bernado, L., Roma, A., Paier, A., Zemen, T., et al.: In-tunnel vehicular radio channel characterization. In: Vehicular Technology Conference, pp. 1–5 (2011)
26. Shivaldova, V., et al: Performance analysis of vehicle-to-vehicle tunnel measurements at 5.9 GHz. In: 30th URSI General Assembly and Scientific Symposium (URSIGASS'11), IEEE, Istanbul, Turkey (2011)
27. Loredo, S., del Castillo, A., et al.: Small-scale fading analysis of the vehicular-to-vehicular channel inside tunnels. Wirel. Commun. Mobile Comput. **2017**(1987437) (2017)
28. Bernado, L., Zemen, T., Tufvesson, F., Molisch, A.F., Mecklenbrauker, C.F.: Delay and doppler spreads of non-stationary vehicular channels for safety relevant scenarios. (2013) CoRR vol. abs/1305.3376
29. Hrovat, A., kandus, G., Javornuc, T.: A survey of radio propagation modeling for tunnels. IEEE Commun. Surv. Tutor. **16**(2), 658–69 (2014)
30. Qureshi, M.A., Noor, R.M., Shamim, A., Shamshirband, S., Choo, K.K.R.: A lightweight radio propagation model for vehicular communication in road tunnels, PloS One **11**(3) (2016)
31. Bilstrup, K., Uhlemann, E., Stroom, E., Bilstrup, U.: On the ability of the 802.11p MAC method and STDMA to support real-time vehicle-to vehicle communication. J. Wireless Commun. Netw. 1–13 (2009)
32. Al Hanbali, A., Altman, E., Nain, P.: A survey of TCP over ad hoc networks. IEEE Commun. Surv. Tutor. **2009**, 22–36 (2005)
33. Torrent-Moreno, M, et al.: Vehicle-to-vehicle communication: fair transmit power control for safety-critical information. IEEE Trans. Veh. Tech. **58**(7) (2009)
34. Mariyasagayam, M.N., Osafune, M., Lenardi, M: Enhanced multi-hop vehicular broadcast (MHVB) for active safety applications. In: 7th IEEE International Conference on ITS Telecommunications (2007)
35. Chehri, A., El Ouahmani, T., Hakem, N.: Mining and IoT-based vehicle ad-hoc network: industry opportunities and innovation. Internet Things 100117 (2019). ISSN: 2542-6605
36. Li, D., Huang, H., Li, X., Li, M., Tang, F.: A distance-based directional broadcast protocol for urban vehicular ad hoc network. In: IEEE (2007)

Real-Time Data Processing in Autonomous Vehicles Based on Distributed Architecture: A Case Study

Yassine El Hafid, Abdessamad El Rharras, Abdellah Chehri, Rachid Saadane, and Mohammed Wahbi

Abstract This work aims to evaluate the real-time processing system in the context of an autonomous vehicle with limited hardware and software capabilities. We elaborate algorithm implemented in 1/10 scale electric car using a line scan camera, speed sensors, and embedded electronic control system. The vehicle navigates in an arbitrary one-lane circuit using an edge detection algorithm. The challenge was to make a complete one loop of the arbitrary circuit in the shortest time with various lighting conditions. The experiments show that several issues were revealed in each step of data sensors processing and need a robust algorithm to handle exceptions caused by multiple disturbances. Furthermore, we paid particular attention to time constraints in embedded processor calculation and actuators response time to achieve reliable critical software control algorithms.

Keywords Autonomous vehicle · Real-time data processing · Embedded system

Y. El Hafid (✉) · A. El Rharras · R. Saadane · M. Wahbi
Laboratory Engineering System, SIRC/LAGeS-EHTP Hassania School of Public Works, Casablanca, Morocco
e-mail: el_hafid_yassine@ehtp.ac.ma

A. El Rharras
e-mail: elgharras.abdessamad@ehtp.ac.ma

R. Saadane
e-mail: rachid.saadane@gmail.com

M. Wahbi
e-mail: wahbi.mo@gmail.com

A. Chehri
Department of Applied Sciences, University of Quebec in Chicoutimi, Chicoutimi, QC G7H 2B1, Canada
e-mail: achehri@uqac.ca

143

1   Introduction

In the past few years, the new era of the autonomous vehicle has rinsing according to the development of new technologies like powerful embedded processor, Lidar, satellite positioning system, MEMS sensors, and lightweight and high power density batteries. However, developing reliable and safe autonomous software need a strong understanding of the environment where the vehicle navigates, especially in the development of the next smart car transportation system where human life is the main concern. The primary issue is the unexpected events that can disturb the behavior of the autonomous.

The development of software capable of handling such complexed environments is possible with using novel methods like ANN [1], SVM [2], deep learning [3], and others. Those methods have an expensive cost time processing and power consumption, especially in an embedded control system [4].

In this article, we expose our studies about a small autonomous vehicle in the scope of our participation in international smart car competition organized by the NXP Semiconductor Company. The goal to achieve is to make one loop of the arbitrary circuit with the maximum speed and without leaving the circuit. Several challenges were placed in the road like hills, chicanes, crosses, and bumps. Furthermore, the lighting conditions were arbitrary by using a different kind of technology, controlling system, intensity, and some time with the parasitic source of light from the external environment.

The next part of this paper is organized as follows. Section 2 presents car kit with different sensors, actuators, and electronic control boards. Section 3 exposes different problematic and issues encountered in the laboratory test. Section 4 describes the proposed embedded software solution and his implementation with time constraints. And finally, a general conclusion is drawn in Sect. 5.

2   Car System Architecture

As we introduced in the previous paragraph, a normalized competition car kit was used for the principle of equal opportunities and fairness. However, the contestant can use additional sensors and actuators with limitations according to competition rules. Figure 1 presents the architecture of our test car used in competition.

2.1   Car Chassis and Actuators

The car chassis as shown in Fig. 2 is the one-piece glass fiber chassis with tow DC motors controlling the rear wheels. For front steering wheels, we use a digital RC
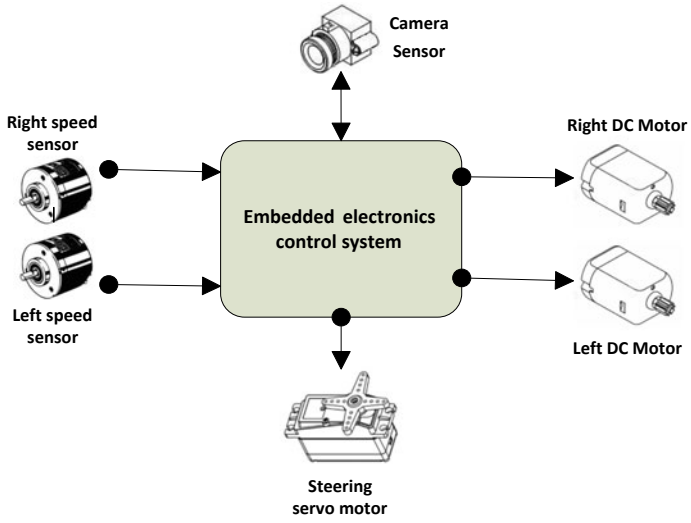
Fig. 1 Car system architecture

Fig. 2 Alamak car model
for NXP CUP competition

servo motor having a torque of 11 kg/cm, which connected by a metallic arm. The overall weight of the car is 750 g (including the batteries).

2.2 Line Scan Camera TSL1401

The primary sensor used was the line scan camera TSL1401, as shown in Fig. 3. It is 128 pixels photodiodes array, which can be controlled by the embedded processor through tow signals CLK and SI (signal integration). The reading value is performed with the ADC module of the microcontroller.

The camera was placed on aluminum road at 30 cm from the floor. Furthermore, the reading point was 60 cm from the center of the front wheels. The lens camera angle is 120° with an infrared filter.

However, the camera has an imperfection in detection. Indeed, we observe that when the camera reads on the white side of the track, especially in the cross. Different level of whites is detected like an arc, as shown in Fig. 4.

Fig. 3 TSL1401 camera module
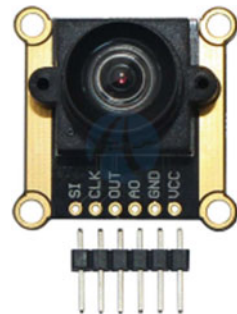
Fig. 4 Line scan camera reading the white part of the track
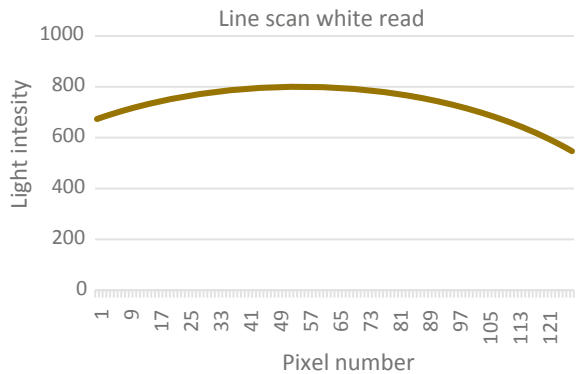
Fig. 5 Incremental
25-mm-dia. Rotary encoder
E6A2-C

Another camera imperfection is the value of the edge on high light intensity. The pixels from 0 to 18 and from 109 to 127 no perform correctly as the pixels in the center of the sensor, so we ignore them in measurement.

2.3 Optical Encoder

As we need to make a speed controller, we implement a system for measuring the speed. We adopted the encoder E6A2-C (incremental 25 mm dia. rotary encoder) as shown in Fig. 5 for providing an electrical pulse signal corresponding to the rotational position and direction of the wheel. In every period T ($T = 10$ ms), we calculate the number of front edges in the entry of one of the pins of the microcontroller. We deduce the angular speed by dividing this number by the period T. To have the linear velocity, we multiply the angular velocity by the radius of the wheel and the transmission ratio of the gear system.

2.4 Electronic Boards Control

We use three-layers stacked boards composed by KL25Z microcontroller clocked at 48 MHz with 128 KB flash program data memory and 16 KB SRAM as the main-board. Power board is based on the BTN7960B motor circuit drivers and interface board for the switch, sensors, display, and Bluetooth communication (Fig. 6).

2.5 Battery

According to rules of NXP CUP competition, only NiMH and Li-ion batteries are allowed for competition with the constraint of 7.2 V as maximum nominal voltage rate and 300 mAh as the maximum capacity. We choose a specific Li-ion battery among various types for the lightweight and higher discharge current needed in car

Fig. 6 KL25Z, power and
interface boards

acceleration. The NiMH batteries are two to three times heavier than Li-ion ones.
Below you will find the characteristics of batteries used in our car model:

- Voltage: Nominal: 3.7 V, maximum: 4.2 V, capacity: 2500 mAh,
- Discharge max: 35 A (pulse), continuous discharge max: 20 A,
- Battery LiNiCoAlO$_2$ (NCA),
- Dimensions: *L*: 65 ± 0.5 mm, *D*: 18 ± 0.5 mm,
- Weight: 44 g by a cell.

3 Car Test Environments and Disturbances Sources

The most challenging aspect of the autonomous vehicle is to be able to handle all
fault conditions. So, the developer needs to understand all the possible sources of
disturbances carefully. In general, several safety tests must be performed in different
situations to trigger fault states and debug embedded software.

3.1 Track Test

For the laboratory test as the NXP CUP competition, we use the same track specifi-
cation according to rules as shown in Fig. 7. The track was made with 3 mm white
plastic sheet with black edges.

The surface was rough enough to keep ground contact of the wheels. However, at
high speed the car slipped and start drift causing error speed measurements. Another
factor like dust can increase this behavior, so we need to keep car wheels and track
clean. On another note, we observe in some cases that a reflective light can be detected
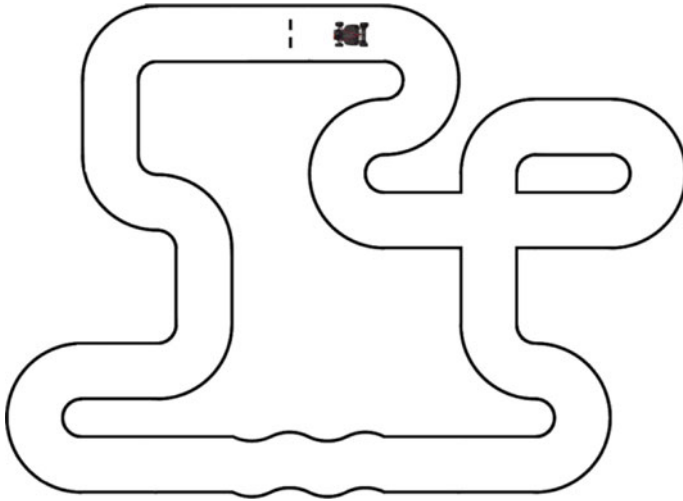by the camera causing the misreading black edges as white.

Fig. 7 Example of car track test

3.2 Lightning System and Parasitic Light Sources

Many lightning technologies were developed in the past few years. In our study, we focus on using the 50/60 Hz light system, which can cause reading fluctuations even if the car is on stop position. Indeed, the integration time of the camera sensor is 10 ms, which is the deadline for our real-time car system. Thus, we developed an adaptive level threshold algorithm to binarize camera data flow. Others technologies like LED or CFL with electronic ballast are easy to handle by the camera, which saw it like DC source of light.

Furthermore, in particular, the situation the car can experiment with glare conditions [5] can resolve by an adaptative exposition time of the photodiodes sensor. Sunlight, powerful projector, or flashlight can cause this issue. In our participation in NXPCUP 2019 final, we had the same problem resolved by putting a filter on the camera lens.

3.3 Mechanical Vibration

Like all vehicles, the mechanical vibrations are generated basically by actuators or unbalanced wheels. In our case, there are two mains issues. The first one is camera misreading [6], which we resolved by detecting and avoiding the data infinite state machine algorithm implemented in our software. The second one is the parasitic values of speed measurements from the two optical encoders [7]. We used a digital low pass filter as a solution.

3.4 Hardware Failure

Numerous hardware failures were observed caused by low-quality spares parts, no well-designed electronics systems, or degradation by car crashes. We took special attention to the wiring and connecting the system to avoid failures during competition.

4 Algorithm Implementation

In this part, we present the architecture of the algorithm with different stages. The data processing flow was achieved in the duration of 10 ms for each camera reading and actuators updating values. We used several programming techniques and dedicated modules in the KL25Z processor to reach our goals.

4.1 Software Architecture

Figure 8 shows the global architecture of our real-time software.

4.2 Camera Data Flow Processing

The first step is camera data acquisition. We use a timer, and ADC interrupts to trigger pixel reading from the camera. The raw data was stocked in an array for the next step processing. After that, we apply a correction function to raw data image, which was modelized as a reverse function of polynomial approximation of camera white reading. Figure 9 shows the curve before and after corrections. In the third step, we binarize data using a dynamic threshold calculated from the max and min of light intensity value at each data reading. After that, we extract the data black and white segment by recording a start pixel segment number and segment length.

4.3 State Machine Car Decision Control Model

As introduced in previous paragraph, the car must run as fast as possible in the track without living it. However, to ensure the safety, we develop a control model using machine state according the car position on the road. In this stage, we took the position and length of the segments to calculate two kinds of parameters. The first one is car position error on the track with fault state avoiding. The adaptive PID
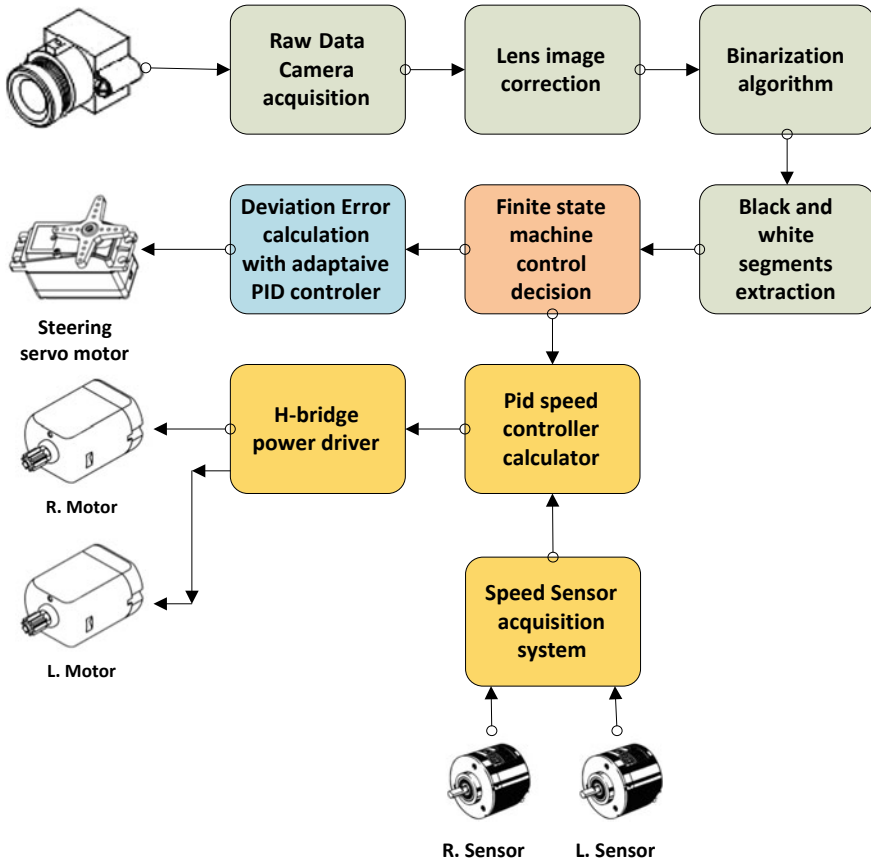
Fig. 8 Architecture of real-time smart car software controller

steering control will use the positioning error. The second one is speed references for each left and right motors used in the PID speed controller. So, we used the simplified control model below:

- **High Speed State**: We use the maximum velocity V_{max} as reference speed for a small value of position error.
- **Low Speed State**: For higher value of deviation error, a bend speed reference V_{rot} is set. A position threshold error P_{eTh} is set to switch between the two states.
- **Fault State**: When the errors occur during camera or sensors reading or for unknown reason. We keep the previous position error and the car does the same calculations as before. In the normal case, we have to stop the car for the safety but in the context of the competition we have to continue until the final line.

Furthermore, the states can detect pattern sequences to start/stop or speed changing according to specials marks on the track (Fig. 10).
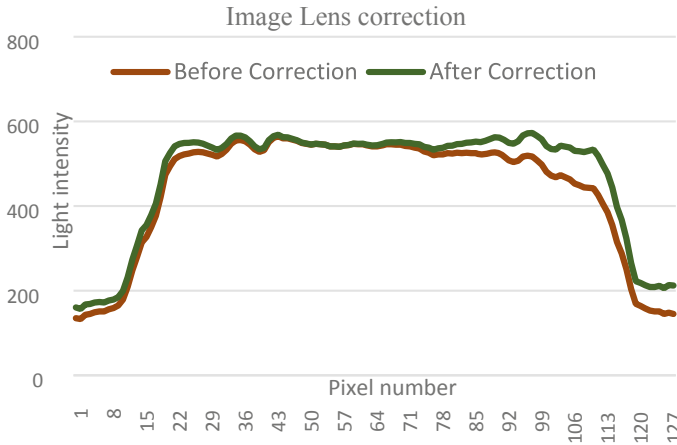
Fig. 9 Raw data camera before and after correction

Fig. 10 Simplified state machine car decision control model
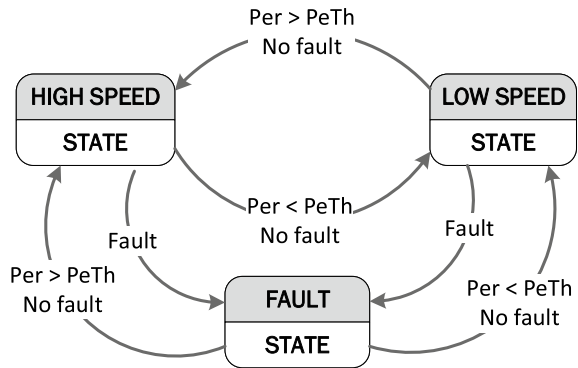
4.4 Speeds Measurement

To achieve speed measurement, we use the sequent measure of each sensor channel by the one-timer module in 5 ms for each sensor. An interrupt help the processor to switch between channels. After that, we apply a digital low pass filter to avoid parasitic values.

4.5 PID Controllers for DC Motors Speed and Adaptative Steering

The PID command consists of three terms *P*, *I*, and *D*, hence, the '*P*' corresponds to the proportional term, '*I*' for the integral, and '*D*' for the term derived from the command. The PID controller is a continuous error calculating the feedback approach. The error calculated as the difference between the reference value and the measured one. It can eliminate the compensation for the state of equilibrium through integral action, and it can anticipate the future through a derivative action.

Those stages aim to implement two PID commands for both DC motors and steering systems with separate excitation (to obtain an adequate response from the process and regulation and have a precise, fast, stable, and robust system). The adjustment of the coefficients (parameters) of the PID command is based on the empirical method of "Ziegler and Nichols" [8].

As the first step, we use the empirical method to specify the coefficient for each DC motor even if they seem the same. We use a wireless data acquisition for recording and get the right factors, which generally depend on the vehicle weight. The steering system is much more complicated because that car behavior depends on the global speed and steering angle. For which we make an array of PID parameters depending on car speed. Thus, the car behaves correctly in curves and straights.

5 Conclusion

In this work, we achieve several improvements in our software system on each stage. We used several methods and techniques to solve different issues caused by a disturbed environment. Excellent skills in hardware and software programming are needed to optimize code implementation. We project for future works to add more sensors like accelerometer and gyroscope to merge data for better control software algorithms, especially for drifting cases. Otherwise, we preconize to use a novel parallel platform like GPU for an embedded system; it seams that will be the better hardware platform for the autonomous vehicles development.

References

1. Katare, D., El-Sharkawy, M.: Embedded system enabled vehicle collision detection: an ANN classifier. In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0284–0289. IEEE (2019)
2. Gao, T., Lai, Z., Mei, Z., Wu, Q.: Hybrid SVM-CNN classification technique for moving targets in automotive FMCW radar system. In: 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6. IEEE (2019)

3. Ren, J., Ren, R., Green, M., Huang, X.: A deep learning method for fault detection of autonomous vehicles. In: 2019 14th International Conference on Computer Science & Education (ICCSE), pp. 749–754. IEEE (2019)
4. Essaid, M., Idoumghar, L., Lepagnot, J., Brévilliers, M.: GPU parallelization strategies for metaheuristics: a survey. Int. J. Parall. Emerg. Distrib. Syst. **34**(5), 497–522 (2019)
5. Hamm, M., Huhn, W.: Glare investigations and safety research on digital light technologies. In: SAE Technical Paper (2019)
6. Verma, M., Collette, C.: Active vibration isolation system for drone cameras. In: Proceedings of International Conference on Vibration Problems: ICOVP (2019)
7. Alejandre, I., Artés, M.: Method for the evaluation of optical encoders performance under vibration. Precis. Eng. **31**(2), 114–121 (2007)
8. Shariff, H.M., Rahiman, M.H.F., Adnan, R., Marzaki, M.H., Tajjudin, M., Jalil, M.H.A.: The PID integrated anti-windup scheme by Ziegler-Nichols tuning for small-scale steam distillation process. In: 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET), pp. 391–395 (2019)

Injury Severity Analysis of Secondary Incidents

Jing Li and Jingqiu Guo

Abstract Limited efforts have been made to unveil the factors affecting the severity of secondary incidents. Compared to primary incidents, secondary incidents are more injury and fatality prone. Secondary incidents that occurred on the Interstate-5 in California within five years were collected. Detailed real-time traffic flow data, geometric characteristics and weather conditions were obtained. First, a random forest-based (RF) feature selection approach was adopted. Then, support vector machine (SVM) models were developed to investigate the effects of contributing factors. For comparison, RF and ordered logistic (OL) models were also built based on the same dataset. It was found that the SVM model has high capacity for solving classification problems with limited data availability. Further, sensitivity analysis assessed the impacts of explanatory variables on the injury severity level. The results can provide guidance for the development of countermeasures and improvement of road safety policies to potentially reduce road trauma caused by secondary incidents.

1 Introduction

Secondary incidents are not rare and relatively injurious on highways, which are defined as incidents occurring within the spatial and temporal boundaries of the impact area caused by a primary incident. The previous research has shown that about

---

This is a short initial version, for more detailed see Li et al. [1]

The original version of this chapter was revised: The belated corrections in the references were incorporated. The correction to this chapter is available at https://doi.org/10.1007/978-981-15-5270-0_19

J. Li · J. Guo (✉)
The Key Laboratory of Road and Traffic Engineering, Ministry of Education,
Tongji University, Shanghai, China
e-mail: guojingqiu@hotmail.com

J. Li
e-mail: jinglion@tongji.edu.cn

155

2.2–3.9% of highway incidents cause secondary incidents [2]. According to the incident data collected during this research, the rate of incapacitating and fatal injuries of secondary incidents is almost twice that of the normal incidents. Therefore, secondary incidents require additional research and road management resources. Understanding the relationships between the injury severity of secondary incidents and their influencing factors can assist road safety management departments to incorporate effective countermeasures and further reduce incident severity levels.

The previous studies have focused on the identification of secondary incidents [3–5] and modeling of the secondary incident prediction (i.e., the risk of secondary incidents, the occurrence time and the frequency of secondary incidents) [6–8]. However, limited efforts have been made on the injury severity analysis of secondary incidents. So far, few studies have investigated the effects of contributing factors on the injury severity of secondary incidents. This could lead to biased conclusions in injury control and safety management and may lead to sub-optimal countermeasures.

Moreover, the previous research of secondary incidents often obtains aggregate traffic data, such as traffic density or annual average daily traffic (AADT), as the independent variables for modeling of secondary incident risk [9]. However, it has been found that real-time traffic data have more accurate capability for measuring traffic conditions than aggregate traffic data [10–12]. Little study has considered the potential influence of real-time traffic flow on the severity level of a secondary incident. Furthermore, ignoring the imbalanced data problem or using a resampling approach to the entire dataset instead of the training set potentially leads to biased models and overfitting estimates [13].

To fill the above-mentioned gaps, this study aimed to investigate the relationship between injury severity of secondary incidents and various factors including traffic flow conditions, weather conditions, primary incident characteristics and geometric design. In order to select significant variables, the random forest (RF) algorithm was used. Support vector machine (SVM) models were developed to investigate the effects of contributing factors on the injury severity of a secondary incident. For comparison, a RF and an ordered logistic (OL) model were also built based on the same dataset. The contributions of this study are as follows:

(1) The study develops a model linking injury severity of secondary incidents with various factors including traffic flow conditions, weather conditions, primary incident characteristics and geometric design.
(2) To address the imbalanced dataset problem, resampling is only performed on the training set instead of the entire dataset. The original, imbalanced validation set is used to assess predictive performance.
(3) Various features are identified that significantly affect the probability of an incident with incapacitating and fatal injuries.

2 Data

Incidents that occurred section (between postmile 138 and 261 km) of north of the Interstate-5 in California from 2010 to 2015 were collected, including the characteristics of incident, real-time traffic flow data, geometric variables and weather data. The characteristics of incidents, traffic flow data and geometric variables were obtained from the Caltrans Performance Measurement System (PeMS). Hourly weather data were collected from the United States National Climatic Data Center (NCDC). A summary of considered variables of geometry, traffic and weather is provided in Table 1.

The speed contour map-based method was used to identify the secondary incidents [14]. First, this method is to distinguish the spatial-temporal impact areas of primary incidents which account for the impacts of recurrent congestions. Second, further incidents were classified as secondary incidents if they occurred in the impact area of the primary incident. In total, 10,012 incidents were extracted and classified as normal ($n = 9443$), primary ($n = 237$) or secondary incident ($n = 332$). Since multiple secondary incidents can be caused by one primary incident, only the injury severity of the first secondary incident was used in this research. Since a few secondary incidents missed injury severity data, a total of 227 data points (i.e., primary incidents and injury severity of corresponding secondary incidents) were selected for further data analysis. Injury severity was categorized using three ordered levels, representing no injury and only property damage (level 1, 81.9%), minor injuries (level 2, 6.6%) and incapacitating and fatal injuries (level 3, 11.5%).

To solve the imbalanced data problem, this study randomly divided the dataset into training, validation and testing sets. Resampling was performed in the training set instead of the entire dataset. The balanced training set was used for model calibration. The validation set was used to tune parameters. Meanwhile, the original, imbalanced testing set was used to assess predictive performance. More specifically, this study uses the Synthetic Minority Oversampling Technique (SMOT) to overcome the problem of imbalanced datasets. 70% of the dataset was used for training and validation purposes and the remaining 30% for testing purposes. Then, the fivefold cross-validation method was used to divide 70% of the dataset into training set and validation set. Meanwhile, to assess the contribution of resampling, a comparison SVM model was developed based on imbalanced training set and tested on the same imbalanced testing set.

3 Modeling

The random forest (RF) algorithm is applied to select significant variables. The SVM is developed to investigate the injury severity analysis. An SVM model defines injury severity prediction as a multi-class classification problem based on the heterogeneous conditions after the primary incident's occurrence. Only small samples (227)

Table 1 Candidate variables for model development

Symbol	Value	Description	Unit
Geometric characteristics			
Width-R	Continuous	Road width	Feet
Width-L	Continuous	Lane width	Feet
Num-L	Discrete	Number of lanes	–
Cn-L	Discrete	Number of changes in Num-L between primary and secondary incident locations	–
Width-O	Binary	1 = outer shoulder width > 12 feet; 0 = otherwise	
On-ramp	Discrete	Number of on-ramps between upstream and downstream loop detectors	–
Off-ramp	Discrete	Number of off-ramps between upstream and downstream loop detectors	–
Diff-ramp	Discrete	Difference between the number of on- and off-ramps	–
HOV	Binary	1 = median lane is a high-occupancy vehicle lane	–
Bottlenecks	Binary	1 = bottleneck section, 0 = otherwise	–
Dura	Continuous	Time difference between the occurrence of primary incident and clearance of the primary incident site	Minute
Peak	Binary	1 = 07:00–09:00,16:30–20:00; 0 = other time	–
Weather characteristics			
Haze	Binary	1 = adverse weather conditions (haze); 0 = otherwise	–
Rain	Binary	1 = adverse weather conditions (rain); 0 = otherwise	–
Visibility	Continuous	Average visibility per hour	Mile
Wind	Continuous	Hourly wind gust speed	mph
Precip	Continuous	Hourly precipitation	mm
Real-time traffic characteristics (measured during a 5–10 min period)			
Avgcnt-up	Continuous	Average vehicle count at the upstream station	Veh/30 s
Avgocc-up	Continuous	Average detector occupancy at the upstream loop detectors	%
Avgspd-up	Continuous	Average vehicle speed at the upstream loop detectors	Mile/h
Devcnt-up	Continuous	Standard deviation of vehicle count at the upstream loop detectors	Veh/30 s
Devocc-up	Continuous	Standard deviation of detector occupancy at the upstream loop detectors	%
Devspd-up	Continuous	Standard deviation of vehicle speed at the upstream loop detectors	Mile/h
Avgcnt-dw	Continuous	Average vehicle count at the downstream loop detectors	Veh/30 s
Avgocc-dw	Continuous	Average detector occupancy at the downstream loop detectors	%

(continued)

Table 1 (continued)

Symbol	Value	Description	Unit
Avgspd-dw	Continuous	Average vehicle speed at the downstream loop detectors	Mile/h
Devcnt-dw	Continuous	Standard deviation of vehicle count at the downstream loop detectors	Veh/30 s
Devocc-dw	Continuous	Standard deviation of detector occupancy at the downstream loop detectors	%
Devspd-dw	Continuous	Standard deviation of vehicle speed at the downstream loop detectors	Mile/h
Difcnt-u&d	Continuous	Count difference between upstream and downstream	Veh/30 s
Difocc-u&d	Continuous	Occupancy difference between upstream and downstream	%
Difspd-u&d	Continuous	Speed difference between upstream and downstream	Mile/h

are available in this study. It is found that the SVM model has high capacity for solving classification problems with limited data availability. For comparison purposes, this study also developed a RF model and a traditional ordered logistic (OL) model based on the same dataset. As compared to the SVM and RF model, the OL model treats the injury severity of secondary incident as an ordinal variable and analyzes the relationship between injury severity and explanatory variables. The major limitation of SVM models is that they cannot identify the impacts of explanatory variables on the dependent variable. For this reason, a variable effect analysis procedure is indispensable. The relationships between explanatory variables and dependent variable were explored in a sensitivity analysis.

3.1 RF Model

In order to select significant variables, Pearson correlation and the RF method were used based on a balanced training set. Highly correlated variables were not included into model simultaneously. Sets including uncorrelated variables were proposed for further variable selection using RF. In this study, the input for RF is represented as Eq. (1), and the output is the injury severity of a single sample. The procedure of RF is sample database selection, decision tree generation and decision tree merger.

$$\{X = [x_{i1}, x_{i2}, \ldots, x_{im}], y_i\} \quad i = 1, 2, \ldots, n; \, y \in \{1, 2, 3\} \quad (1)$$

where X denotes the data samples. n represents the number of data samples. m represents the number of variables in this study. Y represents the injury severity of secondary incident. When a variable was used as a split variable, the RF was used to evaluate the average decrease of the nodes' impurity, which was measured by the

Gini index in this study. A higher value of the Gini index indicated more importance of variables. The out-of-bag (OOB) error rates, measuring the classification error of random forests, were calculated. The lowest OOB was selected to determine the number of trees. In this study, the minimum OOB error rate was observed when the number of trees equals 100. In RF analysis, the Gini index of each variable in each set including uncorrelated variables was calculated, then normalized and ranked from large to small based on the scikit-learn module in Python.

3.2 Kernel Function

The motivation of SVM is to find the best hyperplane that controls the conflict between the demands of classification and maximizing its distance to the nearest data points on each side [15]. More specifically, a binary classification problem is taken as an example to introduce the definition of SVM. The basic form of training data is shown as follows:

$$(x_i, y_i), i = 1, 2, \ldots, l, \ x \in R^n, \ y \in \{\pm 1\} \tag{2}$$

where y_i is the class variable and x_i is the vector composed of n explanatory variables. The hyperplane can be written as follows:

$$(\boldsymbol{\omega} \cdot \boldsymbol{x}) + b = 0 \tag{3}$$

where $\boldsymbol{\omega}$ represents a vector of weights, b represents the displacement and $\cdot$ represents the dot product. To correctly separate the training data from different classes, the maximum-margin hyperplane must meet the following requirements:

$$y_i[(\boldsymbol{\omega} \cdot \boldsymbol{x}) + b] \geq 1 \quad i = 1, 2, \ldots, l \tag{4}$$

The margin can be calculated as $2 / \|\boldsymbol{\omega}\|$. However, the perfect hyperplane may not exist, or it may cause an overfitted model. To solve this problem, SVM finds the hyperplane that not only minimizes model complexity but also minimizes the misclassifications. The optimal classification function can be written as follows:

$$f(\boldsymbol{x}) = \mathrm{sgn}(\boldsymbol{\omega} \cdot \Phi(\boldsymbol{x}) + b) = \mathrm{sgn}\left(\sum_{i=1}^{l} a_i y_i \Phi(x_i) \cdot \Phi(x) + b\right) \tag{5}$$

where $\Phi(x)$ represents feature vector obtained from x which is mapped and a_i are the Lagrange multipliers. In this study, the Gaussian radial basis function (RBF) kernel and sigmoid kernel are both considered. This study used the "one against one" approach to solve the three-level injury severity problem [16, 17].

We minimize the loss function for a selected set of hyperparameters (including the hyperparameters of the kernel function and the hyperparameter for the model complexity and model fit trade-off) to obtain a trained model. This is performed several times for a range of hyperparameters by grid search. The trained models are evaluated based on the performance on the validation set, and the best model is selected based on that. The average performance on validation sets was taken as training performance of the best model.

The OL model has been widely used in fitting structure for an ordinal response. Injury severity of secondary incidents can be defined as ordinal variables that are classified into three levels: 1—no injury and only property damage; 2—minor injury and 3—incapacitating and fatal injuries. Prediction rates of various models were assessed using accuracy, weighted precision and weighted recall. For multi-label classification, precision and recall of each label can be calculated.

3.3 Sensitivity Analysis

SVM for mining data is sometimes regarded as a "black box", where the impacts of explanatory variables on the dependent variable cannot be assessed. Sensitivity analysis is a frequently used method of exploring the relationship between explanatory and dependent variables. Two-stage sensitivity analysis was proposed in the previous SVM incident studies [18, 19] and is also adopted in this study. First, each variable is changed by a pre-specified unit (i.e., delta), while other variables remain unchanged. The detail of delta is shown in Table 2. Second, the SVM model is recalibrated with the altered dataset. Then, comparisons are made between the probabilities of each injury severity level before and after this change. The impacts of each variable are estimated as the percentage change of each injury severity level by a one-unit perturbation of each variable. Sensitivity analysis can be used to estimate the positive and negative impacts of explanatory variables on the dependent variable.
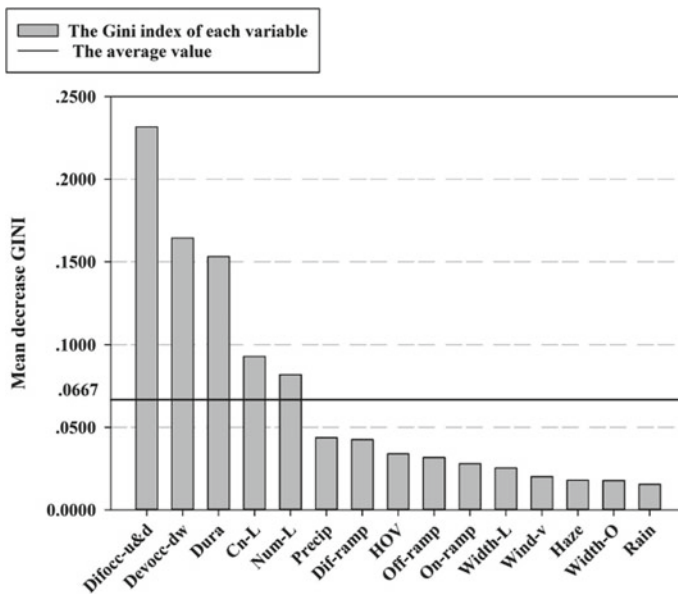
4 Results

The Gini index of the best model set was shown in Fig. 1. In order to choose the most significant variables affecting the injury severity of secondary incidents, the average value for the Gini index of all variables (0.0667) was used as a cutoff value. Five remaining variables were then used to develop the models. A summary of descriptive statistics per variable can be found in Table 2.

Table 2 Variables descriptive statistics and changing unit in sensitivity analysis

Variable	Description	Mean	Std. dev	Max	Delta[a,b]
Difocc-u&d	Occupancy difference between upstream and downstream during 5–10 min period	14.4	12.3	61.1	+1 for M&F-I cases; −1 for No-I cases
DevOcc-dw	Standard deviation of detector occupancy at the downstream loop detector during 5–10 min period	3.3	2.7	18.3	+1 for M&F-I cases; −1 for No-I cases
Dura	Time difference between the occurrence of primary incident and clearance of the primary incident site	41.4	63.3	456.0	+1 for M&F-I cases; −1 for No-I cases
Num-L	Number of lanes	4.6	1.1	6.0	+1 for M&F-I cases; −1 for No-I cases
Cn-L	Number of changes in Num-L between primary and secondary incident locations	2.1	2.6	19.0	+1 for M&F-I cases; −1 for No-I cases

[a]*Note* Delta was used in sensitivity analysis (see Sect. 4.5 for details)
[b]*Note* No-I cases mean no injury and only property damage cases; M & F-I cases mean minor injury cases and fatal injury cases

Fig. 1 Variable ranking of candidate variables based on Gini index

4.1 Model Specifications

The performances of models are shown in Table 3. For the SVM model, the best result is observed when kernel is sigmoid, C is 0.5901, β is 0.6 and θ is 0.0. For the RF model, the best result is observed when the number of trees is 100 (OOB error rate is 16.04%). For the comparison SVM model, the best result is observed when kernel is sigmoid, C is 401, β is 1.6 and θ is 0.0. By comparing the performance of

Table 3 Performance of SVM, RF and OL models

Injury severity[a]	Training (%)	Testing (%)
Sigmoid SVM		
Level 1: No-I	45.2	84.9
Level 2: Minor-I	59.7	33.3
Level 3: Fatal-I	55.3	10.0
Overall accuracy	53.4	69.6
Overall precision	53.7	67.4
Overall recall	53.4	69.6
RF model		
Level 1: No-I	100	81.1
Level 2: Minor-I	100	16.7
Level 3: Fatal-I	100	0
Overall accuracy	100	63.8
Overall precision	100	58.4
Overall recall	100	63.8
OL model		
Level 1: No-I	59.4	47.2
Level 2: Minor-I	14.3	16.6
Level 3: Fatal-I	61.6	40.0
Overall accuracy	45.1	43.5
Overall precision	42.1	61.8
Overall recall	45.1	43.5
Comparison SVM model		
Level 1: No-I	89.6	100
Level 2: Minor-I	4.0	0
Level 3: Fatal-I	28.8	0
Overall accuracy	75.7	87.0
Overall precision	71.2	75.6
Overall recall	75.6	87.0

[a]*Note* No-I means no injury and only property damage; Minor-I means minor injury; Fatal-I means incapacitating and fatal injuries

overall accuracy, precision and recall on testing set, it was revealed that the SVM model with sigmoid kernel performs better than RF and OL models. Moreover, the sigmoid SVM model produced relative superior performance of each level of secondary incident severity on testing set, comparing to RF model and comparison SVM model. Therefore, resampling approach improved the model's performance, and the sigmoid SVM model was preferred and selected as the optimal model for performance discussion in the following and sensitivity analysis in the following section. The overall accuracy and precision of sigmoid model are 69.9 and 67.4%.

As in Table 3, the sigmoid SVM model performs relatively poor with the level 3 category with an accuracy of 10.0%. The previous articles which focus on normal incident severity show consistent results, where the SVM model performs inferior for serious severities, as these severity levels make up only small proportions in the incident database. These studies all report accuracies well under 10% on the validation set for serious injuries.

4.2 *Variable Impact Analysis*

In this study, five input variables were selected for inclusion in the SVM model: Difocc-u&d, Devocc-dw, Dura, Cn-L and Num-L. It can be concluded that higher values of each of the five variables increase the probability of an incident with high injury severity. Meanwhile, the five variables have different impacts on the probability of three injury categories.

(1) For the variable "Difocc-u&d", with changes by delta, the probability of no injury decreases by 4.2%, while the probability of minor injury and fatality category increases by 26.9 and 12.5%. This result indicates that the primary incidents occurring in conditions with a high-occupancy difference between upstream and downstream are more likely to cause secondary incidents with high injury severity.

(2) For the variable "Devocc-dw", when its value is changed by delta, the probability of no injury decreases by 2.1%, while the probability of minor injury category increases by 15.4%. One explanation is that the occurrence of a primary incident may cause more vehicles to switch lanes to avoid the blocked lane. This turbulent traffic condition is likely to propagate up and down in the vehicle platoon, resulting in higher risks of secondary incidents with high injury severity. Therefore, active traffic management interventions such as variable speed limit control can be used to potentially prevent secondary incidents with incapacitating and fatal injuries.

(3) A delta changed in "Dura" leads to 26.9 and 37.5% increases in minor and fatal injury probabilities, respectively. Hence, the duration of the primary incident highly determines the magnitude and degree of congestion. This finding indicates that in order to reduce the injury severity of secondary incidents, fast

incident responses are required such as fast removal of vehicles involved in incidents.

(4) For the variable "Num-L", with changing its value by delta, the probability of minor injury and fatality categories increases by 19.2 and 12.5%. This result indicates that the number of lanes in the occurrence segment of primary incidents is positively associated with injury severity level of secondary incidents. One reasonable explanation is that if an incident happens, more lanes mean drivers can pass the incident location with higher speeds, resulting in higher injury severity in the event of a secondary incident.

(5) Last, for the variable "Cn-L", when changing its value by delta, the probability of fatality category increases by 12.5%. Cn-L respects the complex degree of road segments' geometry. This result indicates that the primary incident occurring in the road segments that frequently change number of lanes is more likely to cause secondary incidents with higher injury severity.

5 Conclusion

This study aimed to investigate the relationship between real-time traffic conditions and injury severity of secondary incidents caused by a primary incident. SVM, RF and OL models were developed to predict the injury severity of secondary incidents, and SVM model outperforms others. The results of feature selection indicated that the variables affecting the injury severity of secondary incidents include: (1) the occupancy difference between upstream and downstream, (2) the standard deviation of detector occupancy at the upstream, (3) the duration of the primary incident, (4) number of changes in lanes' number between primary and secondary incident locations and (5) the number of lanes. The accuracy of the best model on the validation set is found to be 69.6%. However, this study does have small sample size limitations. Additional research is needed to test the robustness and practicability of these study findings to other freeways elsewhere.

References

1. Li, J., Guo, J., Wijnands, J., Yu, R., Xu, C., Stevenson, M.: Assessing injury severity of secondary incidents using support vector machines. Int. J. Transp. Saf. Secur., 1-20 (2020)
2. Xu, C., Xu, S., Wang, C., Li, J.: Investigating the factors affecting secondary crash frequency caused by one primary crash using zero-inflated ordered probit regression. Physica A **524**, 121–129 (2019)
3. Chung, Y.: Identifying primary and secondary crashes from spatiotemporal crash impact analysis. Transp. Res. Rec. **2386**(1), 62–71 (2013)
4. Gupta, V.: Developing an Algorithm to Identify Secondary Incidents. (2018)
5. Kitali, A.E., Alluri, P., Sando, T., Lentz, R.: Impact of primary incident spatiotemporal influence thresholds on the detection of secondary crashes. Transp. Res. Rec. (2019)

6. Goodall, N.J.: Probability of secondary crash occurrence on freeways with the use of private-sector speed data. Transp. Res. Rec. **2635**(1), 11–18 (2017)
7. Jalayer, M., Baratian-Ghorghi, F., Zhou, H.: Identifying and characterizing secondary crashes on the Alabama state highway systems. Adv. Transp. Stud. **37** (2015)
8. Park, H., Haghani, A., Samuel, S., Knodler, M.A.: Real-time prediction and avoidance of secondary crashes under unexpected traffic congestion. Accid. Anal. Prev. **112**, 39–49 (2018)
9. Santos, M.S., Soares, J.P., Abreu, P.H., Araujo, H., Santos, J.: Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]. IEEE Comp. Intell. Mag. **13**(4), 59–76 (2018)
10. Qu, X., Yu, Y., Zhou, M., Lin, C.T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach. Appl. Energy **257**, 114030 (2020)
11. Zhou, M., Yu, Y., Qu, X.: development of an efficient driving strategy for connected and automated vehicles at signalized intersections: a reinforcement learning approach. IEEE Trans. Intell. Transp. Syst. **21**(1), 433–443 (2020)
12. Li, X., Medal, H., Qu, X.: Connected Heterogeneous infrastructure location design under additive service utilities. Transp. Res. Part B **120**, 99–124 (2019)
13. Kitali, A.E., Alluri, P., Sando, T., Haule, H., Kidando, E., Lentz, R.: Likelihood estimation of secondary crashes using Bayesian complementary log-log model. Accid. Anal. Prev. **119**, 58–67 (2018)
14. Xu, C., Liu, P., Yang, B., Wang, W.: Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data. Transp. Res. Part C Emerg. Technolo. **71**, 406–418 (2016)
15. Hastie, T., Tibshirani, R., Friedman, J., Franklin, J.: The elements of statistical learning: data mining, inference and prediction. Math. Intell. **27**(2), 83–85 (2005)
16. Knerr, S., Personnaz, L., Dreyfus, G.: Single-layer learning revisited: a stepwise procedure for building and training a neural network. Neurocomputing 41–50 (1990)
17. Kuang, Y., Qu, X., Wang, S.: A tree-structured crash surrogate measure for freeways. Acci. Anal. Prev. **77**, 137–148 (2015)
18. Chen, C., Zhang, G., Qian, Z., Tarefder, R.A., Tian, Z.: Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accid. Anal. Prev. **90**, 128–139 (2016)
19. Meng, Q., Qu, X.: Estimation of vehicle crash frequencies in road tunnels. Accid. Anal. Prev. **48**, 254–263 (2012)

Communication and Localization Techniques in VANET Network for Intelligent Traffic System in Smart Cities: A Review

Abdellah Chehri, Nordine Quadar, and Rachid Saadane

Abstract The combination of automotive technology and communication networks enables novel systems that improve safety, efficiency, and performance can significantly improve comfort in daily traffic. Vehicle-to-vehicle communication enables new applications through the direct exchange of information between vehicles. In recent decades, this has been intensively researched and standardized technology. The cars thus capture other road users in their environment in smart cities, even beyond visual obstacles. This technology includes digital, wireless communication between vehicles (V2V) or cars and traffic infrastructure (V2I), which is collectively referred to as V2X. V2X communication has a more extended capability range than ultrasonic sensors, cameras, and radars, and can, therefore, alert drivers of dangerous situations earlier and more effectively. Moreover, V2V can be combined with radars and cameras to achieve even greater safety. Vehicle automation and driver assistance systems are also driving forward the promising technology. This paper evaluates state-of-the-art vehicle communication and localization techniques and investigates their applicability on VANET networks for intelligent traffic system.

A. Chehri (✉)
Department of Applied Sciences, University of Quebec in Chicoutimi, Saguenay, QUÉBEC G7H 2B1, Canada
e-mail: achehri@uqac.ca

N. Quadar
University of Ottawa, Ottawa K1N 6N5, Canada
e-mail: nquad016@uottawa.ca

R. Saadane
SIRC2S/LASI EHTP, Route El-Jadida, Casablanca, Morocco
e-mail: rachid.saadane@gmail.com

1   Introduction

The automotive industry is undergoing a significant change with the advent of electrification, autonomous driving, and connectivity. These three factors oblige automotive manufacturers and suppliers to rethink the functioning of the vehicles of the future and their use. The arrival of autonomous and connected electric cars will change our habits and disrupt the transport sector.

The development of the intelligent transportation systems (ITS) was introduced by Japanese Association of Electronic Technology for Automobile Traffic and Driving (JSK) [1]. The main goal was to improve road quality, safety, efficiency, and comfort. This strict and ambitious purpose has driven the researcher's community to look for new techniques, in all aspects of ITS, to fulfil the requirements. As the communication between vehicles presents the pillar to achieve the goal behind intelligent transportation systems, a lot of studies have been conducted to design a new network.

Recent advancement of autonomous vehicle (AV) technologies provides sufficient technical supports and enormous opportunities to control car-following and lane-change manoeuvres.

More advanced AV will use an automated electronic system to replace human drivers. For example, Tesla, a forerunner, provides autopilot function which includes adaptive cruise control, lane-keeping, auto lane change, Auto Park, and automatic emergency steering. These emerging AV technologies have powered individual vehicles with super computation and control capabilities in lane-change location, car-following speed, and acceleration/deceleration, etc., in real time.

One of the fruits of these researchers is the vehicular ad hoc networks (or VANET). VANET can be seen as a particular case of the mobile ad hoc network (MANET) where all nodes move with high speed; these nodes represent the vehicles. This new era of mobile ad hoc networks allows to the nodes (vehicles) the possibility to communicate with each other using vehicle-to-vehicle communication (V2V) and with the infrastructure network of the roadside using vehicle-to-roadside connection (V2R). These two techniques and others are discussed further in this work.

The VANET uses different sensors that are installed inside the car to gather pertinent data or events regarding the environment where the vehicle circulates. This information could be motion, pressure, temperature, sound, vibration, and so on. The used sensors should be able, in addition to collecting these events, to transmit information to the cars connected to the same VANET. This is called inter-vehicle communication [2]. Also, the sensors send data to the driver to provide him/her with traffic and car conditions; this form of communication is called intra-vehicle communication [2].

The vehicular ad hoc network is a particular type of the mobile ad hoc network, as mentioned before. The main difference is that the nodes in the VANET are known and are always the same (vehicles). However, they can be a variety of devices in the case of the MANET. Also, for the VANET, the moving path of the nodes, as

well as the infrastructure, is known. These opportunities can be used to make the connectivity more efficient and the network more useful.

Unlike MANET, the nodes in the VANET are moving with high speed. This fact presents a severe challenge in terms of having an efficient and reliable communication between the nodes [3].

The following sections of this survey are organized as follows. Section 2 discusses the different communication model for transportation application. Section 3 gives an overview of the main VANET application with their specific requirements and discusses the available localization techniques. Section 4 concludes the paper.

2 Communication Model for Transportation Applications

Connected vehicles are vehicles that use any of several different communication technologies to communicate with the driver, other cars on the road (vehicle-to-vehicle [V2V]), roadside infrastructure (vehicle-to-infrastructure [V2I]), and the "Cloud" [V2C]. This technology can be used to not only improve vehicle safety but also to improve vehicle efficiency and commute times. The cars can use one or more of these models to coordinate and send data such as routing information (e.g. direction, speed, position, identity), warning and emergency messages (e.g. traffic update, constructions), and information messages (e.g. service offered in the road like the gas.

Listed below are the types of communication, with links to more information, and some of the benefits of connected vehicles:

2.1 Vehicle-to-Vehicle (V2V)

The idea behind the V2V communication model is to build a virtual bridge between neighbouring cars where information can be transmitted so that the nodes can be the source, destination, or a router. Because of the high speed of the vehicles, this bridge has a short lifetime causing a significant change in the topology of this kind of models. One of the possible solutions is to increase the transmission range of the sensors used in the car, especially in some areas with low penetration ratio. However, the increase in this range is based on transmission power, which is a critical parameter to consider. Also, this increase can lead to interference problems, primarily when many cars circulate in the same area, for example during rush hours. To solve this issue, an attractive solution has been discussed in [4] by adapting the transmission power based on the local density.

Moreover, the privacy issue should not be ignored, as the cars will send personal information. This information could be position, speed, and direction; they can be sent when a vehicle transmits a brief message or in the case when it sends a local event. Some techniques are proposed in [5] to overcome this issue.
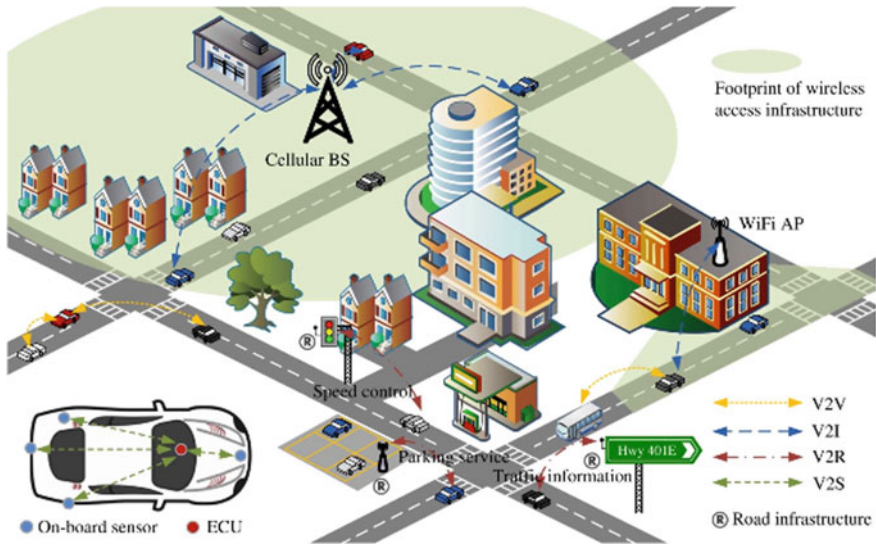
Fig. 1 Example of VANETs communications. http://bbcr.uwaterloo.ca/SubGroup/securitybbcr/vanet.html. [Online] Accessed: 05 March 2020

2.2 Vehicle-to-Infrastructure (V2I)

In the V2I communication model, the vehicles communicate with each other using an intermediate infrastructure that is usually fixed and installed in the roadside, as shown in Fig. 1. This technique is perfect when information should be broadcasted to every connected node in the network [6]. In this case, some problems can be avoided, such as the privacy issue and lifetime connectivity. Usually, a third party manages the intermediate infrastructure, so some security rules can be introduced in order to protect the personal information of the vehicles connecting to the network. But to do that, this third party should be a trusted source, which is not always the case.

Some V2I model can operate without going through an intermediate roadside infrastructure, just a roadside unit is sufficient to manage the connectivity. Also, in this case, increasing the transmission range in some low-density areas could reinforce the connectivity. Again, it is a trade-off between the efficiency and the power consumption.

2.3 Vehicle-to-People (V2P)

The V2P approach encompasses a broad set of road users including people walking, children being pushed in strollers, people using wheelchairs or other mobility devices, passengers embarking and disembarking buses and trains, and people riding bicycles.

2.4 Vehicle-to-Cloud (V2I)

With the emergence of connected infotainment systems (connected navigation, social media, music streaming, and in-car Wi-Fi) and accompanying automotive application frameworks, more advanced vehicle connectivity platforms and cloud capabilities are required. The results in advanced cloud-based connected car platforms with capabilities far exceed those of legacy telematics platforms and also require broadband cellular connectivity, initially 3G, but now increasingly 4G, and with 5G. Additionally, over the air (OTA) is quickly becoming an essential vehicle lifecycle management tool as well as an enabler of analytics and big data approaches. The cloud is quickly becoming the critical technology enabling cars connecting with the wider IoE (or Internet of Everything).

2.5 Vehicle-to-X (V2X)

V2X refers to an intelligent transport system where all vehicles and infrastructure systems are interconnected with each other.

This connectivity will provide more precise knowledge of the traffic situation across the entire road network, which in turn will help to optimize traffic flows, reduce congestion, and cut accident numbers.

2.6 Comparison Between V2V and V2I

Table 1 gives a high-level summary and comparison between these two communication models.

Table 1 Comparison between V2V and V2I

	Advantages	Disadvantages
V2V	Short-range communication Support short message delivery Deployment cost is low Faster communication Low-power consumption	High topology change Long-range communication Privacy and security Short lifetime connectivity Problem of broadcast storm
V2I	Spreading the information widely Support short message delivery Support short message delivery Privacy and security Long lifetime connectivity Low topology change	Deployment cost is high Third party is involved Slower communication Higher-power consumption

3 VANET Application Requirements

As discussed in the previous sections, the ITS main goals are to improve road quality, safety, efficiency, and comfort. The VANET application plays a significant role to achieve this goal by facilitating the communication between the nodes and also providing an efficient transmission in terms of accuracy. Also, the purpose of the VANET technology is to replace the existing application that uses fixed sensors and devices in the road. These traditional applications demonstrate some limitation, especially in terms of maintenance, power supply access, coverage, and related implementation cost.

In this section, the three types of applications are discussed. For each application, the protocols and algorithms that are used to manage the position information are described. Table 2 shows some examples of these applications with their localization requirements [7–13].

Two main types of protocols are applied in this type of application: routing algorithms and data dissemination algorithms. The first type uses the information about the localization to find the next nearer hop to the destination. This information will help the protocol to forward the information about a local or received event to the final destination in the case of hop-by-hop transmission. A standard example of this family called greedy forwarding is discussed in this section. The main goal of the second type is to inform the neighbour vehicle about certain events or services. The vehicle position plays a crucial role to identify the destination. Optimized dissemination of alarm messages (ODAM) protocol is discussed as an example of this type of algorithm in this section.

3.1 Greedy Forwarding Algorithm

As mentioned before, this protocol goal is to transmit the received or generated packet to the next hop. This choice relies on the position estimation of the neighbouring vehicles, hence this information is the key of path optimization of the data to forward.

Table 2 Example of applications for VANET [14–17]

Application	Communication model	Localization accuracy	Range (m)
Estimation of real-time speed	V2V and/or V2I	Medium	100
FCWS: forward collision warning system	V2V and/or V2I	High	100
Effective arterial incident detection	V2I	High	300
Automatic parking	V2V and V2I	High	**
Industrial applications	V2V and V2I	High	**

To illustrate the concept of this algorithm, an example is shown in Fig. 2 [18]. The scenario is as follows: The node x is defined as the event generator, and this event should go through the nodes until it reaches the destination D.

The steps of the algorithm are as follows:

- Search phase ⇒ during this phase, each node broadcasts to its neighbours a beacon about its position, and this data includes also the node's IP address. Neighbours are those nodes located within the transmission range of the host node x.
- Rout Table Building phase ⇒ in this phase, the node builds a table based on the information received from its neighbours; this table contains the ID and the position of all neighbouring nodes. If a node does not receive bacon from its neighbour for a certain time $T = 4.5B$ (with B in the time transmission interval between bacons) [18], the node deletes the node from the table and sends this new information to its neighbour. This can happen in the case where a node leaves the network as the nodes move frequently, or when a node is out of covering range.

The topology of VANET frequently changes because the nodes move and can leave the network at any moment. This fact makes the beaconing phase instable, and the choice of the time transmission interval B becomes crucial. Sending too many beacons in small and repetitive intervals consumes significant energy. To solve this problem, a new strategy can be applied by requesting the beacons from other nodes only when there is traffic to forward. This can save energy and decrease the cost related to this communication.

One of the inconvenience of the greedy forwarding is when the forwarding path drawn by this protocol takes a longer distance to the destination than the direct transmission. This specific topology is shown in Fig. 3; the destination D is located just after the transmission range of the host node x. Using the forwarding protocol, in this case, is not efficient.

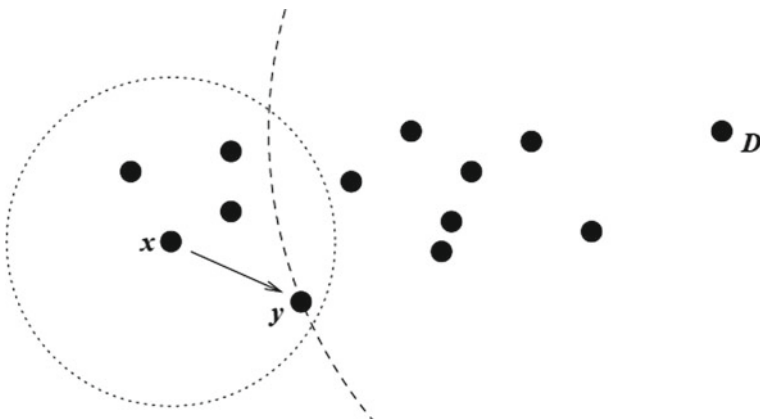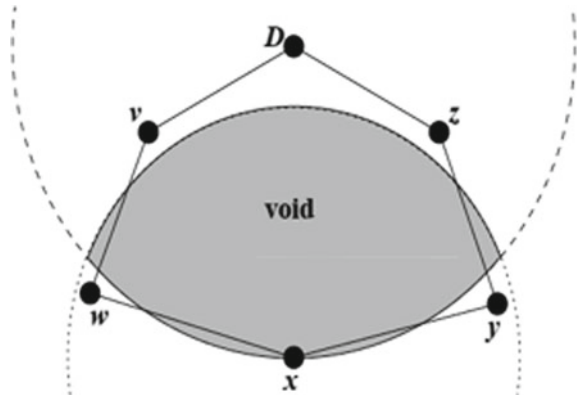**Fig. 2** Greedy forwarding algorithm

Fig. 3 Greedy forwarding
limitation

3.2   *Optimized Dissemination of Alarm Messages (ODAM)*

The ODAM protocol goal is to disseminate an alarm message within the VANET in order to inform other connected vehicle about a critical event happening in the front side of the platoon. ODAM uses information such as the position and direction of the affected vehicle to give precise alarm message, so it needs the driver direction in addition to its location. The alarm messages can be an obstacle, accident, etc.

Figure 4 shows an example of ODAM use [19], where x is the accident vehicle that wants to inform the rest of the nodes about the dangerous event. As shown, not all nodes are within the transmission range of x. So, the protocol uses the neighbour vehicles (are called relay) to rebroadcast the event and so on until the event reaches all concerned nodes.

The algorithm goal is to forward the message about the accident happening to the vehicle (x) to all concerned nodes. The involved nodes can be in both directions, more particularly those driving towards the event. For this reason, it is essential to include the direction of the accident vehicle. To optimize the number of transmissions message, and by consequence of the power consumption, the protocol uses a defer time. A transitory "wait and see" time slot is used if there is no one behind them in the same transmission range. Below are the steps for this algorithm:

- (x) broadcast an alarm message to all vehicle within its transmission range, in this case (a) and (b).
- (a) and (b) will wait for a particular time before rebroadcasting the event; this time called the defer time. In this case, the defer the time of (a) should be higher than (b) as the distance between (a) and (x) is smaller than between (b) and (x). So the defer time is proportional to the distance to the accident vehicle.
- If the node (for instance (a)) does not receive any rebroadcasting from other nodes, it concludes that there is no vehicle behind within the transmission range of (x). In this case, it decides to rebroadcast the event in order to reach other nodes.
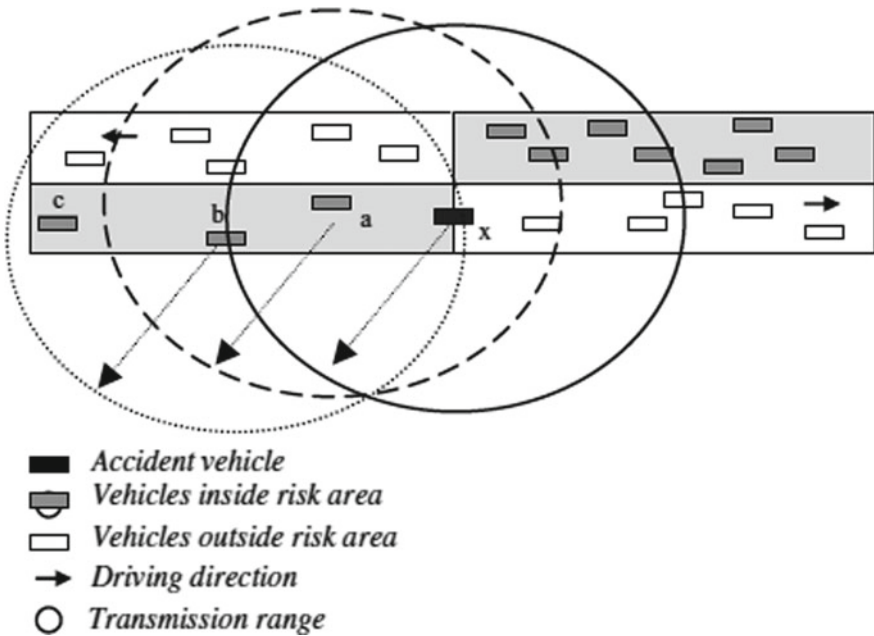
Fig. 4 Greedy forwarding limitation [19]

- If within the defer time it receives a rebroadcasting, it just ignores the event because it understands that there is someone else behind that will take care of the retransmission.

3.3 MAP-based Localization

The last discussed technique of this part is the map-based localization. This technique is used in the navigation applications to enhance the efficiency of the VANET and provide reliable services to assist the driver and increase the safety by giving an estimated localization of the car in the map. By knowing some information about the road using the map, an automatic control application can be feasible and more efficient. For instance, the automatic driver can be enhanced if it has information about the position of the car in the road, the distance to the lane limit, coming intersections.

Table 2 summarizes some examples of VANET applications with the main requirements based on what was discussed in the previous sections.

4   Conclusion and Road Map for Future

Vehicular ad hoc networks (VANETs) have been gaining significant attention from the research community due to their increasing importance for building an intelligent transportation system. However, the implementation is far more complicated than most people realize. It involves many aspects including technology, wireless spectrum, many governmental regulatory bodies, and cooperation of automakers and adaptation of the technology as a whole. This paper presented a comprehensive study of various most significant communication techniques for VANET. Different requirements for ITS application were discussed and classified into three different groups: accurate, inaccurate, and high-accuracy location-aware applications. Each method has its specific use with advantages and drawbacks. So, it is hard to say that there is a perfect method that can be used anytime and anywhere. The goal behind the VANET is to provide the driver with continuous information about the traffic conditions.

Further research will be required to validate and refine the performance and robustness of the most promising localization techniques. Moreover, the integration of cooperative technologies requires new study to find the optimal way to fuse these localization techniques to achieve more reliability and accuracy enhancement.

References

1. Rawat, D.B., Yan, G.: Infrastructures in vehicular communications: status, challenges and perspectives. In: M. Watfa (eds.) IGI Global (2010)
2. Chehri, A., Mouftah, H.T.: Autonomous Vehicles in the Sustainable Cities, the Beginning of a Green Adventure. Sustainable Cities and Society (SCS), vol. 51. Elsevier (2019)
3. Zhou, M., et al.: Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: a reinforcement learning approach. IEEE Trans. Intell. Transp. Syst. **21**(1), 433–443 (2020)
4. Zhou, M., Qu, X., Jin, S.: On the impact of cooperative autonomous vehicles in improving freeway merging: a modified intelligent driver model-based approach. IEEE Trans. Intell. Transp. Syst. **18**(6), 1422–1428 (2017)
5. Rawat, D.B., Yan, G., Popescu, D., Weigle, M., Olariu, S.: Dynamic Adaptation of Joint Transmission Power and Contention Window in VANET. VTC (2009)
6. Parno, B., Perrig, A.: Challenges in securing vehicular networks. Paper presented at the Fourth Workshop on Hot Topics in Networks (HotNets-IV)
7. Dornbush, S., Joshi, A.: Street smart traffic: discovering and disseminating auto- mobile congestion using VANETs, In: Proceedings of the IEEE VTC, Spring, pp. 11–15 (2007, April)
8. Michaelian, M., Browand, F.: Field experiments demonstrate fuel savings for close-following, California PATH Program, Institute of Transportation Studies. University of California at Berkeley (2000)
9. Tatchikou, R., Biswas, S., Dion, F.: Cooperative vehicle collision avoidance using inter-vehicle packet forwarding. In: Proceedings of Global Telecommunications Conference (2005)
10. Shah, R., Wolisz, A., Rabaey, J.: On the performance of geographical routing in the presence of localization errors ad hoc network applications. In: IEEE International Conference on Communications, vol. 5, pp. 2979–2985 (2005

11. Karp, B., Kung, H.T.: GPSR: greedy perimeter stateless routing for wireless sensor networks. In: Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'00), Boston (2000)
12. Benslimane, A.: Localization in vehicular ad hoc networks. In: Proceedings Systems Communications, pp 19–25 (2005)
13. Dia, H., Thomas, K.: Development and evaluation of arterial incident detection models using fusion of simulated probe vehicle and loop detector data. Inf. Fusion **12**, 20–27 (2011)
14. Chehri, A., El Ouhamani, T., Hakem, N.: Mining and IoT-Based Vehicle Ad-hoc NETwork: Industry Opportunities and Innovation. Internet of Things: Engineering Cyber Physical Human Systems, Elsevier, pp. 100117. ISSN 2542-6605 (2019)
15. Chehri, A., Quadar, N., Saadane, R.: Survey on localization methods for autonomous vehicles in smart cities. In: 4th International Conference SCA (2019)
16. Chehri, A., Mouftah, H.: Localization for vehicular ad hoc network and autonomous vehicles, are we done yet? In: Connected and Autonomous Vehicles in Smart Cities. CRC Press, Taylor & Francis Group (2020)
17. Chehri, A., Chehri, H., Hakem, N., Saadane, R.: Realistic 5.9 GHz DSRC vehicle-to-vehicle wireless communication protocols for cooperative collision warning in underground mining. In: 13th International Conference on Human Centred Intelligent Systems (STS-20), Croatia (2020)
18. Bachmann, C., Roorda, M.J., Abdulhai, B., Moshiri, B.: Fusing a Bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation. J. Intell. Transp. Syst. **17**, 152–164 (2013)
19. Tsai, M.-F., Chilamkurti, N., Ho, P.-F., Lu, Y.-C.: Implementation of dedicated short range communications combined with radar detection for forward collision warning system. J. Wirel. Netw. Broadband Technol. **2**, 49–63 (2012)

Modelling the Relationships Between Headway and Speed in Saturation Flow of Signalised Intersections

Yang Teng, Jin Xu, Kun Gao, and Ziling Zeng

Abstract The headways between vehicles in the traffic flow of intersections are one of the crucial variables for reasonable signal timing setting and intersection configuration design. Many studies apply constant discharge headways to calculate the saturation flow rate, and scarce studies quantitatively investigate the relationship of headway and speed in the saturation flow. This study endeavours to model the headway–speed relationships of saturation traffic flow at the signalised intersection. Five typical intersections with large traffic demand in Golden Coast City are surveyed to collect data regarding vehicles' discharging speed and headways. The least squared method and the fitting degree test are applied to model the headway–speed relationships at the signalised intersections and compare the models' fitting performance. The results indicate that the headway is significantly associated with speed. The headway increases with decreasing speed crossing the intersections. The empirically and quantitatively calibrated relationships between speed and headway can be used to calculate the saturation flow rate in the intersections with different discharging speeds and further support the design of intersections with large traffic demand.

Keywords Saturation traffic flow · Headway · Discharging speed · Model regression

Y. Teng
School of Civil and Environmental Engineering, University of Technology Sydney, Sydney 2007, Australia
e-mail: avaterbta@gmail.com

J. Xu
36-3-401, He Tian Shang Cheng Liu He Road, Binjiang District, Hangzhou, Zhejiang, China

K. Gao (✉) · Z. Zeng
Architecture and Civil Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden

1   Introduction

With the fast development of road networks in big metropolises, the roles of signalised intersections have been increasingly important for transport systems. The operational efficiency of signalised intersections greatly impacts the performance of the whole transport system in terms of traffic efficiency [1] and traffic safety [2]. The signal timing is a vital component in the traffic design of signalised intersections. Reasonable design of signal timing is helpful to increase the traffic capacity and reduce the traffic conflicts at signalised intersections. Discharging headway is one of the crucial parameters in the signal timing design [3]. Lost time and the saturation flow can be calculated using the value of discharge headway. The discharge headway is defined as the time interval between two successive vehicles in the same driveway passing the stop lines at signalised intersections. Generally, the queue discharge headway becomes constant after the fourth or fifth vehicles [4]. The lost time is determined by adding up the differences between each vehicle's actual headway and their estimated saturation headway [5]. Therefore, the saturation headway is important for designing signal timing of signalised intersections [6, 7].

As for the saturation flow rate, it is defined as the peak value of queue discharge during the green time and calculated by the equation $S = 3600/h$ [8] where S represents the saturation flow rate and h denotes the saturation headway. In traditional methods, the discharge headway is regarded as a constant in calculating the saturation flow rate, which may lead to incorrect estimation of the saturation headway [9]. There are several factors affecting the estimation of saturation headway such as the time variance, negligence of capacity factors and inappropriate equations. Consequently, incorrect estimation of saturation flow rate will result in incorrect delay prediction and signal timing [10].

Many studies have addressed the methods for measuring the saturation flow rate [11]. Teply et al. [12] put forward a method of splitting green time into five or ten seconds as a survey period and counting the vehicles passing the intersections during each period for completing the statistical analysis of saturation flow rate. However, there are various numbers of vehicles counted in each short periods of green time, which lead to unreliable data collection. Highway Capacity Manual [13] proposed a method for calculating the saturation flow rate at the signalised intersection. In this method, the value of saturation headway is approximated as the constant mean value of the headways of the vehicles passing signalised intersection from the fourth vehicle (in order to minimise the impact of vehicles' start-up lost time) to the last vehicle. However, the actual headways might be unsaturated using this method. For example, traffic demand is unable to arrive at the stop bar, no more than four vehicles in the same queue and a large percentage of heavy vehicles in the cycles [14]. Some other opposing viewpoints against regarding saturation headways as a constant value [15]. According to Rahman et al. [8], treating the discharge headway as a constant would result in underestimation or overestimation of the saturation flow rate in special cases. Li and Prevedouros [16] used the traditional methods to compute the saturated headway and concluded that the traditional site investigation

of saturated flow by applying the first 12 vehicles could result in underestimation of the flow rate of left-turn vehicles and overestimation of the flow rate of straight-going vehicles. Shao et al. [17] conducted another investigation and show that the mean headway in China was 2.04 s by applying traditional methods. They reported that the root-mean-square deviation of the conventional method was between 0.4 and 0.9 s indicating great fluctuation. In brief, measuring the saturation headway as a constant is not correct and may result in biases in estimating the saturation flow rate in the signalised intersections.

As for modelling the relationship between speed and traffic volume, Greenberg [18] proposed a logarithmic model to describe the relationship between speed and volume when the traffic volume is large. Underwood [19] put forward an exponential model to depict the speed–volume relationship when the traffic volume is small. Another typical speed–volume model was the Greenshields model, which was derived from the linear relationship between speed and density. The Highway Capacity Manual [13] adopted this model. However, there was scarce research in modelling the saturation headway–speed relationship. The modelling of saturation headway–speed is to quantitatively measure the functional relationship between saturation headway and speed. Through the quantitative function, the headways can be attained from the observed travelling speed and be used to calculate the saturation flow rate at the signalised intersections.

In summary, two research gaps are found based on the above literature review: one is that existing research treats the discharge headway as a constant and lacks of investigating the variation of saturation headway; another gap is that there is scarce research about saturation headway–speed modelling. This study stands in the wake of the two research gaps to quantitatively and empirically model the headway–speed relationship in saturation flow of signalised intersections based on observed traffic data in field surveys. It is to shed light on the relationship of time headway and driving speed in the intersections and support the accurate estimation of the saturation flow rate of intersections for the scientific design of intersections in terms of signal timing setting and configuration.

2 Methods

2.1 Traffic Survey Design Ad Data Collection

In Fig. 1, the black line denotes the driving distance of vehicles. The red lines (i.e. the position of head of vehicles) and the blue lines (i.e. the stop lines in the existing direction) represent the starting point and ending point of vehicles passing the intersections, respectively. Traffic investigators started timing using an electric timer once the green light turned on (i.e. $t = 0$) and ended the timer once the last vehicle in the queue went through the blue line. The time when the first car started to move was recorded and used as the reaction time (t_0). The reaction time denotes the driver

Fig. 1 Example of the
investigated intersections:
Markeri St and Bermuda St
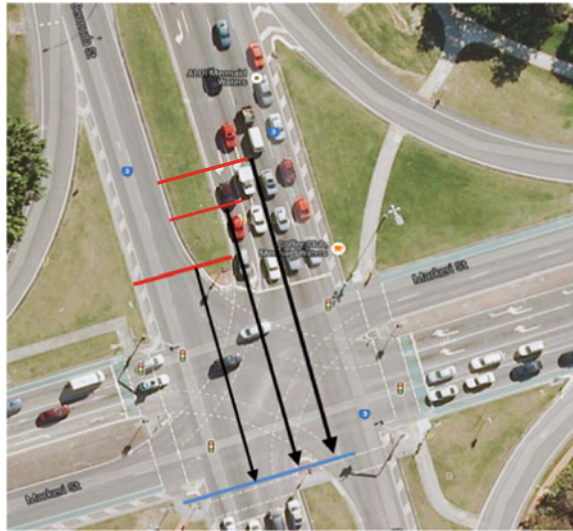
of the first vehicle starts to go after the green light goes on. The t_1 is measured as
the time when the first car crossed the stop line. The time of the following vehicle i
passing the blue line was successively recorded and labelled as t_i ($i = 1, 2, 3, …, n$).
Five intersections in the Gold Coast City were surveyed to cover wide geographic
situations. The details of the investigated intersections are shown in Table 1. All of
the selected intersections have big traffic flow during peak time. The road conditions
of these intersections have standard road widths and flat gradient.

The headways can be computed by the time interval of two successive cars passing
the blue line. As for speed, there are two kinds of speed for measurement: time mean
speed and space mean speed. The time mean speed is defined as the arithmetic
average velocity of the vehicle passing a certain point. Space mean speed is defined
as the harmonic average velocity of vehicles passing a certain road. The space mean

Table 1 Summary of selected intersections

Intersections	Fairway Dr and Nerang Broadbeach Rd	Slatyer Ave and Bundall Rd	Crestwood Dr and Olsen Ave	Parklands Dr and Olsen Ave	Markeri St and Bermuda St
The recording time	1 h	1 h	1 h	1 h	1 h
Cycle numbers	27	31	18	24	32
Mean vehicle numbers per cycle	10.5	10.4	8.3	8.8	8.2

speed is used in this study as per the Highway Capacity Manual [13]. The driving distance of each vehicle is calculated by the observed vehicle length and the gaps between two successive vehicles and can be denoted as the equation:

$$S = L + 8 * n \tag{1}$$

where L is the length of the intersection (from the stop line to the opposite stop line), and n represents the index of vehicles. The space mean speed of the vehicle n is given by

$$v = (L + 8 * n) / (t_{n+1} - t_0) \tag{2}$$

where t_0 is the reaction time (Fig. 2).

2.2 Modelling Methods

The least squares method is used to model the headway–speed relationship using the collected data. The headway and speed are the dependent variable (i.e. headway, denoted as y) and the explanatory variable (i.e. speed, denoted as x), respectively. The best fitting polynomial functions could be generalised by the basic best-fitted model. The variable types could be simply transferred to get the straight lines even if the variables are exponent, exponentiation or logarithm [20].

In order to test the fitting degree of the proposed models, two fitting measurements are used: R-squared and root-mean-square error. R-squared is also called "the ratio of determination", which reflected the explanation percentage of independent variables' all variation caused by the dependent variables [21]. For example, if the R-squared is 0.8, it indicates that the regression relationship can explain the 80% of the variation of the independent variable. In other words, if the dependent variable can be controlled to be invariable, the variation level of independent variable can be reduced to 20%.

The value of R-squared is determined by the sum of the square of regression (SSR) and the sum of the difference (SSD). The SSR is the sum of the difference between predicted variable $\hat{y}$ and the mean of the observed variable $\bar{y}_i$ and is given by

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2 \tag{3}$$

where SST is the sum of the difference between observed variable y_i and the mean of the observed variable $\bar{y}_i$ and is given by

$$SST = \sum_{i=1}^{n} (y_i - y_i)^2 \tag{4}$$
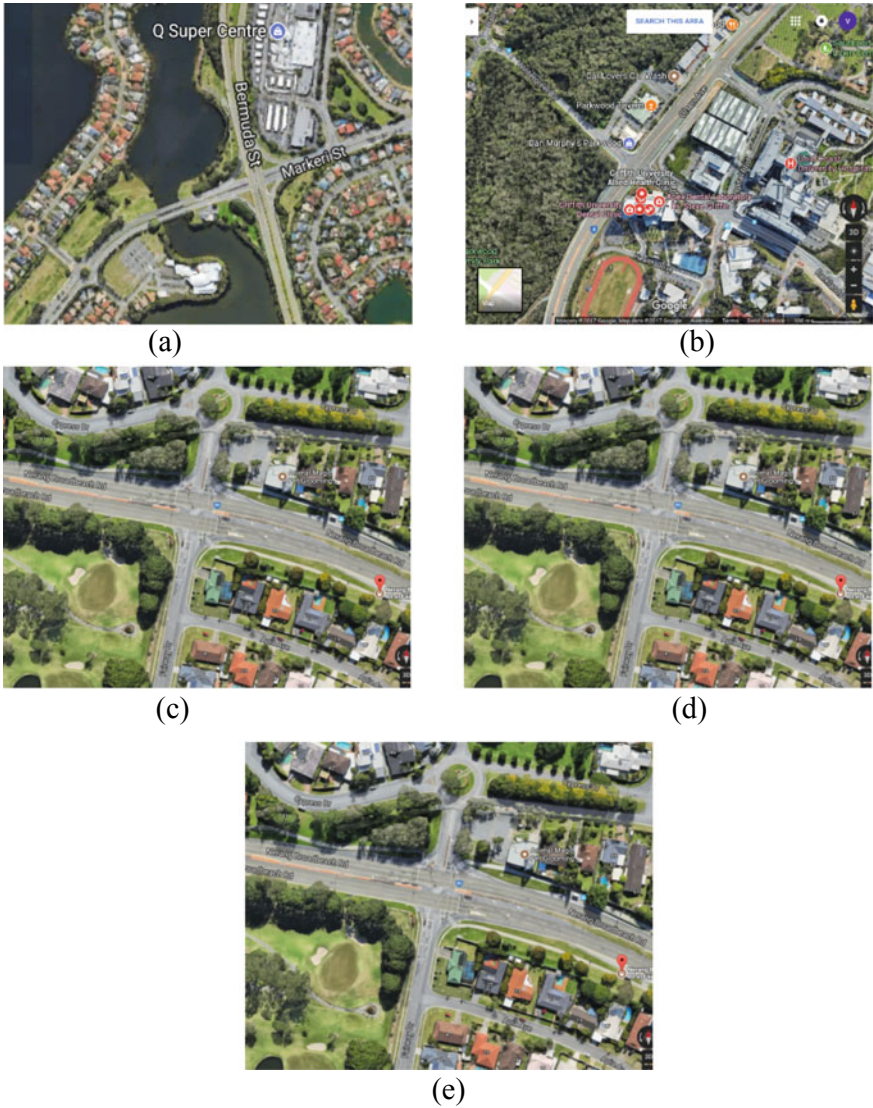
Fig. 2 Investigated intersections: **a** intersection of Markeri Street and Bermuda Street, **b** intersection of Oslen Avenue and Parklands Drive, **c** Fairway Drive and Nerang Broadbeach Road, **d** Fairway Drive and Nerang Broadbeach Road and **e** Crestwood Drive and Olsen Avenue

R-squared can be calculated by

$$R\text{-squared} = \frac{\text{SSR}}{\text{SST}} \tag{5}$$

The value of R-squared is between 0 and 1. When the value of R-squared closes to 1, it shows that the most variation of the independent variable in the fitted model is influenced by the variation of the dependent variable, reflecting high goodness of fit. R-squared closes to 0 means low goodness of fit [22].

The root-mean-square error (RMSE) is a commonly applied measure for measuring the difference value between the predicted and observed variables. Its value is usually the predicted value of a model. The RMSE is standard error for the difference between the forecasting value and the observed value [23], and is given by

$$\text{RMSE} = \sqrt{\frac{\sum d_i^2}{n}} \tag{5}$$

where d_i is the difference between the observed value y_i and predicted value $\hat{y}_i$, and n is the number of observed points. The less the value is, the higher the measurement accuracy is.

3   Result and Analysis

For the sake of model fitness and result presentative, the data form cycles with more than eight vehicles are selected to analyse. The cycles with R-squared over 0.5 are mainly analysed and shown in this section. The rest cycles that have R-squared with less than 0.5 value are all rejected and not analysed.

For the intersection of Markeri St and Bermuda St, cycle 20 and cycle 24 are selected as per their models' R-squared, and corresponding results are shown in Fig. 3. Both models based on the data from cycle 20 and cycle 24 have R-squared larger than 0.5. However, RMSE in cycle 20 is 0.197, and RMSE in cycle 24 is 0.687, so RMSE in cycle 20 is smaller than that in cycle 24. The model of cycle 20 has a comparatively better fitting degree for this intersection. The equation of the fitted model using cycle 20 is $y = -0.6041 * x + 6.69$. The slope is negative. It demonstrates that the saturation headway will decrease when speed increases at this
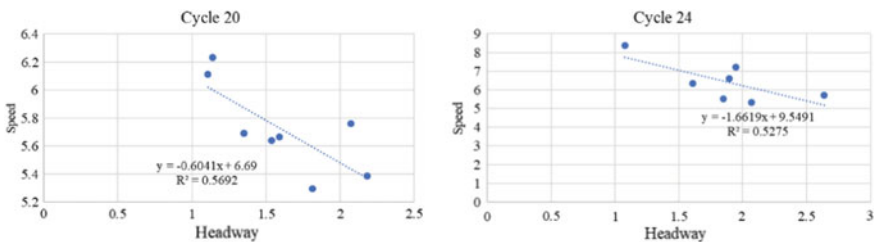

Fig. 3 Result of the intersection of Markeri St and Bermuda St. The y-axis is the speed (m/s) in second, and the x-axis is headway (s)

intersection. Since the saturation flow rate is the inverse of saturation headway, so it will increase with speed improvements.

For the intersection of Slatyer Ave and Bundall Rd, data from all the cycles are analysed. However, the results are not representative for analysis and interpretation due to its large variance. All the calibrated models have R-squared less than 0.5. So they are not used for further discussion.

For the intersection of Fairway Dr and Nerang Broadbeach Rd, cycle 6 and cycle 13 are selected as per their models' R-squared, and corresponding results are shown in Fig. 4. R-squared of the models based on data of cycle 6 and 13 are 0.6614 and 0.5169, respectively. As for RMSE, RMSE of cycle 6 is 0.430, which is bigger than the RMSE of cycle 13 of 0.173. The equations of the fitted models are $y = -1.1741 * x + 8.1695$ and $y = -0.1466 * x + 4.9561$ for cycles 6 and 13, respectively. The calibrated slope of cycles 6 and 13 is less than 0, implying the headway decreases with the mean space speed.

For the intersection of Parklands Dr and Olsen Ave, cycle 15 and cycle 24 are selected, and the results are shown in Fig. 5. The R-squared in cycle 15 is 0.5219, and R-squared in cycle 24 is 0.518. As for RMSE, RMSE for cycle 15 is 0.639, and RMSE for cycle 24 is 0.502. Again, we can observe that the headway decreases with the increase of mean space speed.
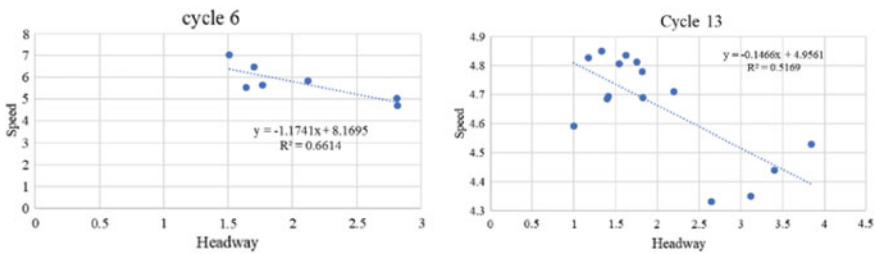

Fig. 4 Result of the intersection of Fairway Dr and Nerang Broadbeach Rd. The y-axis is the speed (m/s) in second, and the x-axis is headway (s). Cycle 6 means the data from sixth single cycle
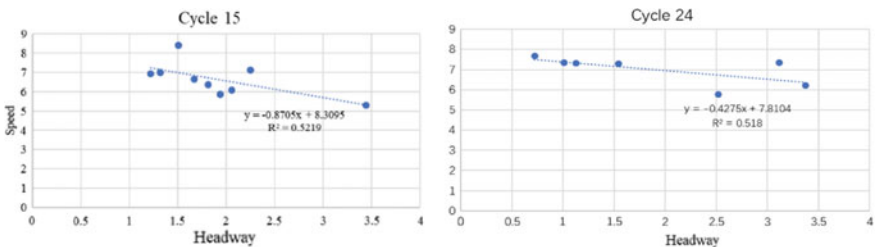
Fig. 5 Result of the intersection of Parklands Dr & Olsen Ave. The y-axis is the speed (m/s) in second and the x-axis is headway (s)

Fig. 6 Result of the intersection of Crestwood Rd and Olsen Ave. The *y*-axis is the speed (m/s) in second, and the *x*-axis is headway (s)
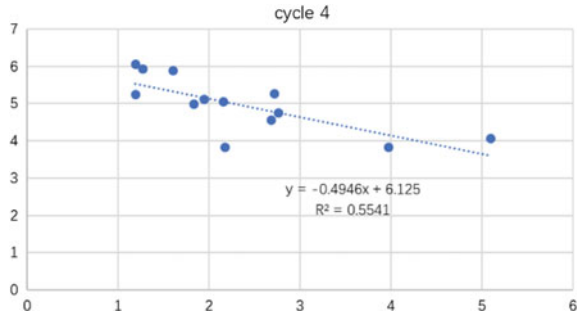
For the intersection of Crestwood Rd and Olsen Ave, cycle 4 is selected because its model's *R*-squared is more than 0.5, and the result is shown as Fig. 6. The *R*-squared in cycle 4 is 0.5541, and the RMSE is 0.582. The equation of the fitted model is $y = -0.4946*x + 6.125$. The slope of the model is negative. Therefore, the saturation headway will decrease when speed increases at this intersection. Because the saturation flow rate is the inverse of saturation headway, so it will increase when speed increases.

4 Conclusions

The existing research commonly uses a constant discharge headway to calculate the saturation flow rate in the signalised intersections, and scarce studies investigate the relationship of headway and speed in the saturation flow. This study quantitatively and empirically analyses the headway–speed relationship in saturation flow of signalised intersections. Five typical intersections with large traffic demand in Golden Coast City are surveyed to collect data regarding vehicles' discharging speed and headways. The least square method and the fitting degree test are applied to establish the headway–speed models at the signalised intersections and compare the models' fitting performance. The models with best fitting are identified based on the values of *R*-squared and RMSE. The results indicate that the headway is significantly related to the mean space speed. The headway increases with decreasing speed crossing the intersections. The empirically and quantitatively calibrated relationships between speed and headway can be used to calculate the saturation flow rate in the intersections with different discharging speeds and further support the design of intersection timing setting to fulfil the traffic demand and reducing traffic delays.

References

1. Qu, X., Zhang, J., Wang, S.: On the stochastic fundamental diagram for freeway traffic: model development, analytical properties, validation, and extensive applications. Transp. Res. Part B **104**, 256–271 (2016)
2. Xu, C., Yang, Y., Jin, S., Qu, Z., Hou, L.: Potential risk and its influencing factors for separated bicycle paths. Accid. Anal. Prev. **87**, 59–67 (2016). https://doi.org/10.1016/j.aap.2015.11.014
3. Zhou, M., Yu, Y., Qu, X.: Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: a reinforcement learning approach. IEEE Trans. Intell. Transp. Syst. (2019)
4. Roess, R.P., Prassas, E.S., McShane, W.R.: Traffic engineering. Pearson/Prentice Hall (2004)
5. Minh, C.C., Sano, K.: Analysis of motorcycle effects to saturation flow rate at signalized intersection in developing countries. J. Eastern Asia Soc. Transp. Stud. **5**, 1211–1222 (2003)
6. Li, X., Ghiasi, A., Xu, Z., Qu, X.: A piecewise trajectory optimization model for connected automated vehicles: exact optimization algorithm and queue propagation analysis. Transp. Res. Part B: Methodol. **118**, 429–456 (2018)
7. Zhang, J., Qu, X., Wang, S.: Reproducible generation of experimental data sample for calibrating traffic flow fundamental diagram. Transp. Res. Part A: Policy Pract. **111**, 41–52 (2018)
8. Akcelik, R.: Traffic Signals: Capacity and Timing Analysis. ARRB, Australia (1981)
9. Rahman, M.M., Nur-Ud-Deen, S., Hassan, T.: Comparison of saturation flow rate at signalized intersections in Yokohama and Dhaka. Proc. Eastern Asia Soc. Transp. Stud. **5**, 959–966 (2005)
10. Tarko, A.P., Tracz, M.: Uncertainty in saturation flow predictions. Red **1**, P2 (2000)
11. Qu, X., Wang, S., Zhang, J.: On the fundamental diagram for freeway traffic: a novel calibration approach for single-regime models. Transp. Res. Part B: Methodol. **73**, 91–102 (2015)
12. Teply, S., Allingham, D. I., Richardson, D. B., Stephenson, B.W.: Committee on the Development of the Canadian Urban Transportation Capacity Guide, Canadian Capacity Guide for Signalized Intersection, Canada (2008)
13. Transportation Research Board: Highway Capacity Manual. Transportation Research Board, National Research Council, Washington, DC, USA (2010)
14. Koonce, P., Rodegerdts, L.: Traffic signal timing manual (No. FHWA-HOP-08-024). United States. Federal Highway Administration (2008)
15. Ren, L., Qu, X., Guan, H., Easa, S., Oh, E.: Evaluation of roundabout capacity models: an empirical case study. J. Transp. Eng. ASCE **142**(12), 04016066 (2016)
16. Li, H., Prevedouros, P.: Detailed observations of saturation headways and start-up lost times. Transp. Res. Record: J. Transp. Res. Board, **1802**, 44–53 ((2002))
17. Shao, C.Q., Rong, J., Liu, X.M.: Study on the saturation flow rate and its influence factors at signalized intersections in China. Procedia-Soc. Behav. Sci. **16**, 504–514 (2011)
18. Greenberg, H.: An analysis of traffic flow. Oper. Res. **7**(1), 79–85 (1959)
19. Underwood, R.T.: Speed, volume, and density relationships. In: Quality and Theory of Traffic Flow. Bureau of Highway Traffic. Yale University, New Haven, CT (1961)
20. Abdi, H.: The method of least squares, encyclopedia of measurement and statistics. CA, USA: Thousand Oaks (2007)
21. Draper, N.R. Smith, H.: Applied Regression Analysis, vol. 326. Wiley (1998)
22. Cottrell, A., 2003. Regression analysis: basic concepts, Wake Forest University, Department of Economics, ECN, 215
23. Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. Int. J. Forecast. **22**(4), 679–688 (2006)

Shore Power Price Competition Between Ports

Jingwen Qi, Shuaian Wang, and Xiaobo Qu

Abstract Air pollution and climate change arouse consistent attention of international community. Shipping industry, being one of the most important transport methods, carries more than 80% of the total international trade and has been recognized as a potential source of air pollutant mitigation. In order to reduce emissions of marine traffic, especially in the area of coastal waters, regulations about the quality of marine fuel have been carried out, and the maximum sulphur content allowed for marine fuel becomes increasingly stringent as time goes by. In order to comply with the regulations, shipping has to take various measures, including adopting electric power from shore while berthing. Shore-side electricity, also called cold ironing, refers to the use of electricity from shore side while berthing at the port instead of auxiliary engine. In recent years, shore power has been adopted in an increasing number of ports; in China, most ports are able to provide shore power for ships while berthing. For ships with shore power facilities, the price of shore-side electricity is an element that can influence their choice of port to visit. It is an incentive for ports to lower the power price. This paper tends to investigate what is the best price to maximize the port's total benefit in the competition with other ports in the same group. In order to describe the competition among ports, game theory is applied, and the Bertrand model is adopted.

Keywords Shore power price · Bertrand model · Game theory

J. Qi (✉) · S. Wang
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong 999077, China
e-mail: jingwen.qi@connect.polyu.hk

S. Wang
e-mail: hans.wang@connect.polyu.hk

X. Qu
Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg 412 96, Sweden
e-mail: xiaobo@chalmers.se

1 Introduction

The problem of air pollution and climate change arouse consistent attention of international community. Series of conventions and agreements have been signed among countries and regions such as the Kyoto Protocol and the Delhi Declaration [5]. Shipping industry, being one of the most important transport methods, carries more than 80% of the total international trade [4]; UNCTAD [6, 8]. Recently, shipping sector has been recognized as a potential source of air pollutant mitigation. Relative researches (IMO 2014, 2009) show that the NOx emission from marine traffic makes up 15% of global anthropogenic emission, SO_2 emitted by ships constitutes 13% of the total anthropogenic sources, as for CO_2 marine traffic accounts for approximately 2.7% of the annual emissions. Although the proportions are relatively low, 70% of the marine traffic exhausted emissions are emitted in the area within 400 km of coastline, and have bad influence for air quality, ecological environment and public health of coastal cities. For instance, in Shanghai, 11% of NOx, 12.4% of SO_2 and 5.6% of PM are emitted at the area of the port of Shanghai [2]. At the same time, over 95% of ship currently in use adopt diesel engines for propulsion, and for the cost reason, these engines always burn low-quality fuel oil with high sulphur content. According to the comments to the NSW Environment Protection Authority [7], the fuels used in marine traffic are on average 2700 times dirtier than those used by land vehicles.

In order to reduce emissions of marine traffic, especially in the area of coastal waters, regulations about the quality of marine fuel have been carried out, and the maximum sulphur content allowed for marine fuel becomes increasingly stringent as time goes by. After 1 January 2020, as required by IMO, all the ships should use fuel oil with sulphur content no higher than 0.5%, for ships sailing in the designated emission control areas 0.1% would be the upper limit of sulphur content.

As one of the most important parties in shipping industry, shipping companies can contribute to the reduction of marine traffic emissions. Shore-side electricity, also called cold ironing, refers to the use of electricity from shore side while berthing at the port instead of auxiliary engine. To make use of shore-side electricity system, both the port and the ship need to install some extra equipment. To supply shore-side electricity, port needs to expand the capacity of its substation and install extra socket box and cable operator. Ships also need to be equipped with some special devices to use shore power. Both the port and the shipping company have to make essential extra investment before take advantage of shore power. However, using shore-side electricity can reduce SO_2, NOx and other PM emissions by up to 90% (the comments to the NSW Environment Protection Authority [7]), therefore, produce environmental benefit, enhance the local and regional air quality, and contribute to public health. So, the government will provide subsidies to promote the use of shore power. And with reasonable price, port could even earn some profit from selling the power to ships. In recent years, shore power has been adopted in an increasing number of ports; in China, most ports are able to provide shore power for ships while berthing [2]. For shipping companies, since the installation of shore power system is a one-time investment but can bring long-term revenue in the future it is worthwhile to equip the

ships. As increasing number of ports adopt shore power system and the regulation of fuel oil sulphur content becomes more stringent, the earnings of equipping ships with shore power facilities will be more remarkable. The proportion of ships adopt shore-side power will grow in the future.

A port group consists of several ports that are approximate in geographic location. Therefore, these ports have similar transportation network and partly overlapping economic hinterland. With the development of port capacity and the collection and distribution efficiency, internal competition of port group becomes fiercer. For ships with shore power facilities, the price of shore-side electricity is an element that can influence their choice of port to visit. The ships attracted by preferential power price bring the same port charge revenue and the cargo handling fee in the long term, at the same time, cause much less air pollution than traditional ships. It is an incentive for ports to lower the power price. This paper tends to investigate what is the best price to maximize the port's total benefit in the competition with other ports in the same group. In order to describe the competition among ports, game theory is applied.

2   Literature Review

Shore power system, also referred to as cold ironing, is a promising method to reduce the ship emissions while berthing. According to IMO (2012), shore-side power is "a measure to improve air quality in ports and port cities, to reduce emissions of air pollutants and noise and, to a lesser extent, to reduce carbon dioxide through ships at berth replacing on-board generated power from diesel auxiliary engines with electricity supplied by the shore." Through the use of shore power, the reliance on auxiliary engines to power ships on berthing can be eliminated. The environmental and social benefits are hot topics in research of shore power system, an essential number of existing studies focus on them. Generally speaking, relative studies come up to similar conclusion that the application of shore power system can reduce the air pollutants and mitigate the environmental burden. Various investigations have been conducted to examine the effect of shore power system at different areas and ports. Vaishnav et al. [9] compared the economic costs of shore power system and the benefit it can bring at US ports. Result shows that when quarter to two-third ships call at US ports use shore power instead of auxiliary engines while berthing, the environmental benefits will exceed the economic loss of shipping companies for retrofitting their ships, and have a surplus in the region of $70–150 million. Winkel et al. [10] investigated situations in European ports, and found that through applying shore power system, 800,000 tons' carbon emissions would be reduced by the year 2020, and generate about €2.94 billion public health benefits. According to Ballini and Bozzo [1], with 60% ships using shore power the system will bring €2.97 million external health savings for the port of Copenhagen. For the port of Aberdeen, the environmental benefits are evaluated to be around £1.3 million [3].

Game theory is a new branch of modern mathematics and an important part of operational research. There is no precise and uniform definition for game theory,

but according to John C. Harsanyi, the winner of the 1994s Nobel Prize in Economics, game theory is a theory of strategic interaction. Born in 1928 when Von Neumann proved the basic principle of game theory, game theory becomes one of the most extensively used analytical tools in economics, and has been applied in many disciplines.

3 Model

According to different criteria, games can be divided into several types. In general, all the games can be identified as cooperative game or non-cooperative game, based on the whether there are binding agreements among parties in the game. From the perspective of behaviour time sequence, games are further divided into dynamic game and static game. As the terms suggest, in a static game, all the participants make their decisions at the same time or they do not decide simultaneously but the later actor is blind about the action taken by the former actors. In a dynamic game, participant move in a time sequence and they are aware of the action of competitors that move before them. Based on the understanding level of participants on their competitors, games are divided into two categories, complete information game and incomplete information game. There are some classical models in game theory, such as Cournot model, Bertrand model, Hotelling model and Stackelberg model. The Bertrand model is used to describe the price competition between two suppliers. In the Bertrand model, competitors decide their prices of homogeneous products, and the price will influence the sales volume and market share, it is consistent with the fact that the shore power provided by different ports are the same. In order to avoid the Bertrand paradox, three methods are adopted: first is to consider capacity constraints, the second is to remove the assumption of product homogeneity and the third is to introduce dynamic factors into the model. In this model, the shipping company preference for ports is considered in the competition. For ports in the same group, although they are near, but their economic hinterlands, cargo volumes and service content are not identical, so liner shipping routes have their own preferences in port choosing. So in this paper, the Bertrand model with customer preference is adopted to analyse the optimal shore power pricing strategy for ports.

3.1 Assumptions

Before building the mathematical model, some assumptions are made to characterize the problem that is investigated.

(i)   There are two ports in a port group competing in the price of shore power. Before the providing of shore power, both of them have steady ship customer base.

(ii)   Ports in a port group provide service to liner shipping companies for the same price. Namely, for ships visiting different ports, the port charge is assumed to be the same.

(iii)   Adopting shore power can reduce all of the exhaust emissions of berthing ships.

(iv)   For all the ports, the numbers of ships that are equipped with shore power facility are known before the decision of port side shore power facility installation and pricing.

(v)   For each port, the number of visiting ships with shore power facilities for each port before the shore power pricing movement is certain and known.

(vi)   All ships with on-board shore power facility will choose to use it while berthing.

(vii)   Shipping companies prefer ports that they choose to visit originally. This preference results from advantages of ports such as geographic location and service level. So changing port will bring some loss. In that case, only when the economic profit, namely the cost savings of shifting to another port, is higher than a threshold will the ship route changes it is visiting from port i to port j.

(viii) The fuels used by each ship while berthing are identical, at the same time, ships consume same amount of fuel oil for each hour berthing at the port. On the one hand, the economic gain for a ship to switching from fuel oil to shore power is proportional to the berthing time. Also, the environment cost for ships to use fuel oil is also proportional to berthing time.

(ix)   The capacities of ports for traditional ships and ships using shore power are assumed to be adequate. As a result, ships using shore power will not compete with traditional ships for berth.

(x)   Berthing times of ships visiting both ports obey the uniform distribution in the same interval.

3.2   The Bertrand Model

Before the mathematical model, we explain the parameters and sets.
    Sets and parameters:

Port: The set of ports in the port group.
P: The port charge for both ports (USD/hour).
EnC: The environmental cost of the port area when 1 tonnage fuel oil is consumed by berthing ship (USD/ton).
E: The electricity that is generated by consuming 1 tonnage fuel oil (kwh/ton)
PF: The power fee the port pays to the power company (USD/kwh).
S_i: The number of ships adopting shore power at port i before providing shore power. $i \in 1, 2$.

S'_i: The number of ships adopting shore power at port i after providing shore power. $i \in 1, 2$.

T_{ij}: The loss of shipping company to change from port i to port j. $i, j \in 1, 2, i \neq j$.

BT^i_{min}/BT^i_{max}: The minimum/maximum berthing time of ships at port i. $i \in 1, 2$.

Bunker: The bunker price of fuel oil used by ships at berth (USD/ton).

Fuel: The mass of fuel oil used by ships at berth for each hour (ton/hour).

Decision variables:

Power P_i: The price of shore power at port i.

Ports in the group make pricing strategy of shore power to maximize its own profit $Pro_i, i = 1, 2$. It is assumed that the port should gain some profit by providing shore power to ships. In another word,

$$(\text{Power } P_i - E) + \frac{\text{EnC}}{E} > 0, i = 1, 2 \tag{1}$$

Otherwise, the port will not choose to install the shore power infrastructure and provide electricity to ships, and the pricing problem will not exist. Also, we assume that the using shore power at berth can reduce the cost of ships compared with using fuel oil. Namely

$$\text{Bunker} > E \times \text{Power } P_i, i = 1, 2 \tag{2}$$

Otherwise, no ships will choose to using electricity from shore. Equations (1) and (2) determine the upper and lower limit for the power price Power $P_i \in \left(\frac{\text{EnC}}{E} + \text{PF}, \frac{\text{Bunker}}{E} \right)$. It is assumed in this paper that $\frac{\text{EnC}}{E} + \text{PF} < \frac{\text{Bunker}}{E}$.

3.3 Profit Function

Power $P_1 >$ Power P_2. For the situation that Power $P_1 >$ Power P_2, all the shore power ships visiting port 2 will keep their choice of port. For ships using shore power and visiting port 1, they may change their choice for the economic profit from lower shore power price in port 2. From simple deduction, we know that with the same shore power prices of both ports Power $P_1 >$ Power P_2 ship with longer berthing time will get more cost saving in shore power. So first we calculate the price with which the utility of the ship with the longest berthing time to switch from port 1 to port 2 equals to 0.

$$(\text{Power } P_1 - \text{Power } P_2) \times BT^1_{max} - T_{12} = 0 \tag{3}$$

From Eq. (3), we deduct the value of Power $P_1 = \frac{T_{12}}{BT^1_{max}} + \text{Power } P_2$. Then, we calculate the price with which the utility of the ship with the shortest berthing time to switch from port 1 to port 2 equals to 0.

$$(\text{Power } P_1 - \text{Power } P_2) \times \text{BT}^1_{\min} - T_{12} = 0 \tag{4}$$

From Eq. (4), we deduct the value of Power $P_1 = \frac{T_{12}}{\text{BT}^1_{\min}} + \text{Power } P_2$.

When Power $P_1 \leq \frac{T_{12}}{\text{BT}^1_{\max}} + \text{Power } P_2$, no ship will change the visiting port so the ship number visiting each port remain the same, $S'_1 = S_1$, $S'_2 = S_2$. The profit of two ports are

$$\text{Pro}_i = [(\text{Power } P_i - \text{PF})E + \text{EnC}] \times \text{Fuel}$$
$$\times \frac{S_i}{BT^i_{\max} - BT^i_{\min}} \int_{BT^i_{\min}}^{BT^i_{\max}} x \mathrm{d}x, \ i = 1, 2. \tag{5}$$

The total profit consists of the profit of selling electricity to ships and the reduced environmental cost.

When $\frac{T_{12}}{\text{BT}^1_{\max}} + \text{Power } P_2 < \text{Power } P_1 \leq \frac{T_{12}}{\text{BT}^1_{\min}} + \text{Power } P_2$, assume the ship with berthing time BT' satisfies the following equation

$$(\text{Power } P_1 - \text{Power } P_2) \times \text{BT}' - T_{12} = 0. \tag{6}$$

We can calculate the value of $\text{BT}' = \frac{T_{12}}{\text{Power } P_1 - \text{Power } P_2}$. Ships that berth at port 1 longer than BT' will change their choice and visit port 2. So we have

$$\begin{cases} S'_1 = S_1 - S_1 \times \frac{BT^1_{\max} - BT'}{BT^1_{\max} - BT^1_{\min}} \\ S'_2 = S_2 + S_1 \times \frac{BT^1_{\max} - BT'}{BT^1_{\max} - BT^1_{\min}}. \end{cases} \tag{7}$$

The profit of two ports are like the following equation:

$$\begin{cases} \text{Pro}_1 = [(\text{Power } P_1 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S_1}{BT^1_{\max} - BT^1_{\min}} \int_{BT^1_{\min}}^{BT'} x \mathrm{d}x \\ \quad - [P - \text{EnC} \times \text{Fuel}] \frac{S_1}{BT^1_{\max} - BT^1_{\min}} \int_{BT'}^{BT^1_{\max}} x \mathrm{d}x \\ \text{Pro}_2 = [(\text{Power } P_2 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S_2}{BT^2_{\max} - BT^2_{\min}} \int_{BT^2_{\min}}^{BT^2_{\max}} x \mathrm{d}x \\ \quad + [(\text{Power } P_1 - \text{PF}) \times E \times \text{Fuel} + P] \frac{S_1}{BT^1_{\max} - BT^1_{\min}} \int_{BT'}^{BT^1_{\max}} x \mathrm{d}x \end{cases} \tag{8}$$

Port 1 loses the revenue result from the ships that switch from port 1 to port 2 but gain the electricity selling profit and environmental cost savings for ships that choose to stay. The profit of port 2 consists of two parts, the saved environmental

cost and profit of selling the shore power to ships that originally visit port 2 and the revenue bring by the ships that are attracted form port 1.

When Power $P_1 > \frac{T_{12}}{BT_{\min}^1} + $ Power P_2, all the ships with shore power facility visiting port 1 will change their decision and choose to visit port 2. So $S_1' = 0$, $S_2' = S_2 + S_1$. The profits of two ports are:

$$
\begin{cases}
\text{Pro}_1 = -[P - \text{EnC} \times \text{Fuel}] \frac{S_1}{BT_{\max}^1 - BT_{\min}^1} \int\limits_{BT_{\min}^1}^{BT_{\max}^1} x \, \mathrm{d}x \\[2em]
\text{Pro}_2 = [(\text{Power } P_2 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S_2}{BT_{\max}^2 - BT_{\min}^2} \int\limits_{BT_{\min}^2}^{BT_{\max}^2} x \, \mathrm{d}x \\[2em]
\qquad + [(\text{Power } P_1 - \text{PF}) \times E \times \text{Fuel} + P] \frac{S_1}{BT_{\max}^1 - BT_{\min}^1} \int\limits_{BT_{\min}^1}^{BT_{\max}^1} x \, \mathrm{d}x
\end{cases}
\tag{9}
$$

Port 1 loses the revenue result from all the ships with shore power facility so its profit is negative. The profit of port 2 consists of two parts, the saved environmental cost and profit of selling the shore power to ships that originally visit port 2 and the revenue bring by the ships that are attracted form port 1. Since the profit of port in this situation is negative so it must not be the Nash equilibrium of the game.

Power $P_1 <$ Power P_2. Following the same logic in the above section, we can deduce the profit function for the two ports when Power $P_1 <$ Power P_2.

$$
\text{Pro}_1 =
\begin{cases}
[(\text{Power } P_1 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S_1}{BT_{\max}^1 - BT_{\min}^1} \int\limits_{BT_{\min}^1}^{BT_{\max}^1} x \, \mathrm{d}x, \\[1.5em]
\quad \text{if Power } P_2 \leq \frac{T_{21}}{BT_{\max}^2} + \text{Power } P_1 \\[2em]
[(\text{Power } P_1 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S_1}{BT_{\max}^1 - BT_{\min}^1} \int\limits_{BT_{\min}^1}^{BT_{\max}^1} x \, \mathrm{d}x \\[1.5em]
\quad - [(\text{Power } P_2 - \text{PF}) \times E \times \text{Fuel} + P] \frac{S_2}{BT_{\max}^2 - BT_{\min}^2} \int\limits_{BT'}^{BT_{\max}^2} x \, \mathrm{d}x, \\[1.5em]
\quad \text{if} \frac{T_{21}}{BT_{\max}^2} + \text{Power } P_1 < \text{Power } P_2 \leq \frac{T_{21}}{BT_{\min}^2} + \text{Power } P_1
\end{cases}
\tag{10}
$$

$$\text{Pro}_2 = \begin{cases} [(\text{Power } P_2 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S_2}{\text{BT}_{max}^2 - \text{BT}_{min}^2} \int\limits_{\text{BT}_{min}^2}^{\text{BT}_{max}^2} x\,\mathrm{d}x, \\[2mm] \quad \text{if Power } P_2 \le \frac{T_{21}}{\text{BT}_{max}^2} + \text{Power } P_1 \\[3mm] [(\text{Power } P_2 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S_2}{\text{BT}_{max}^2 - \text{BT}_{min}^2} \int\limits_{\text{BT}_{min}^2}^{\text{BT}'} x\,\mathrm{d}x \\[3mm] \quad - [P - \text{EnC} \times \text{Fuel}] \frac{S_2}{\text{BT}_{max}^2 - \text{BT}_{min}^2} \int\limits_{\text{BT}'}^{\text{BT}_{max}^2} x\,\mathrm{d}x, \\[3mm] \quad \text{if } \frac{T_{21}}{\text{BT}_{max}^2} + \text{Power } P_1 < \text{Power } P_2 \le \frac{T_{21}}{\text{BT}_{min}^2} + \text{Power } P_1 \end{cases} \quad (11)$$

In which $\text{BT}' = \frac{T_{21}}{\text{Power } P_2 - \text{Power } P_1}$.

3.4 Nash Equilibrium

For simplicity, we assume that $S_1 = S_2$, $T_{21} = T_{12}$, $\text{BT}_{max}^1 = \text{BT}_{max}^2$ and $\text{BT}_{min}^1 = \text{BT}_{min}^2$. So in the following subsection, we use S to represent S_1 and S_2, T to represent T_{21} and T_{12}, BT_{max} to represent BT_{max}^1 and BT_{max}^2, BT_{min} to represent BT_{min}^1 and BT_{min}^2.

After reorganizing, we have the profit function for port 1 under different situations:

$$\text{Pro}_1 = \begin{cases} [(\text{Power } P_1 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S}{\text{BT}_{max} - \text{BT}_{min}} \int\limits_{\text{BT}_{min}}^{\text{BT}_{max}} x\,\mathrm{d}x, \\[2mm] \quad \text{if Power } P_2 - \frac{T}{\text{BT}_{max}} \le \text{Power } P_1 \le \text{Power } P_2 + \frac{T}{\text{BT}_{max}} \\[3mm] [(\text{Power } P_1 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S}{\text{BT}_{max} - \text{BT}_{min}} \int\limits_{\text{BT}_{min}}^{\text{BT}_{max}} x\,\mathrm{d}x - \\[3mm] [(\text{Power } P_2 - \text{PF}) \times E \times \text{Fuel} + P] \frac{S}{\text{BT}_{max} - \text{BT}_{min}} \int\limits_{\text{BT}''}^{\text{BT}_{max}} x\,\mathrm{d}x, \\[3mm] \quad \text{if Power } P_2 - \frac{T}{\text{BT}_{max}} > \text{Power } P_1 \ge \text{Power } P_2 - \frac{T}{\text{BT}_{min}} \\[3mm] [(\text{Power } P_1 - \text{PF}) \times E + \text{EnC}] \times \text{Fuel} \times \frac{S}{\text{BT}_{max} - \text{BT}_{min}} \int\limits_{\text{BT}_{min}}^{\text{BT}'} x\,\mathrm{d}x \\[3mm] \quad - [P - \text{EnC} \times \text{Fuel}] \frac{S}{\text{BT}_{max} - \text{BT}_{min}} \int\limits_{\text{BT}'}^{\text{BT}_{max}^1} x\,\mathrm{d}x, \\[3mm] \quad \text{if } \frac{T}{\text{BT}_{max}} + \text{Power } P_2 < \text{Power } P_1 \le \frac{T}{\text{BT}_{min}} + \text{Power } P_2 \end{cases}$$

$$(12)$$

In Eq. (12), we have $\text{BT}' = \frac{T}{\text{Power } P_1 - \text{Power } P_2}$, $\text{BT}'' = \frac{T}{\text{Power } P_2 - \text{Power } P_1}$. And the demand functions of port 1 and port 2 are symmetric. In order to find the Nash equilibrium, we take the partial derivative of the profit function.

$$\frac{\partial \text{Pro}_i}{\partial \text{Power } P_i} = 0, i = 1, 2 \qquad (13)$$

$$\frac{\partial \text{Pro}_1}{\partial x} = \begin{cases} \frac{1}{2} E \times \text{Fuel} \times a \times \left(\text{BT}_{\max}^2 - \text{BT}_{\min}^2\right), \text{if } y - \dfrac{T}{\text{BT}_{\max}} \leq x \leq y \\ \quad + \dfrac{T}{\text{BT}_{\max}} \\ \frac{1}{2} E \times \text{Fuel} \times a \times \left(\text{BT}_{\max}^2 - \text{BT}_{\min}^2\right) + [(y - \text{PF}) \times E \times \text{Fuel} + P] \\ \quad \times a \times T^2 \left(\dfrac{1}{y - x}\right)^3, \\ \text{if } y - \frac{T}{\text{BT}_{\max}} > x \geq y - \frac{T}{\text{BT}_{\min}} \\ \frac{1}{2}\text{Fuel} \times a \times \left\{ E \times \left[\left(\dfrac{T}{x - y}\right)^2 - \text{BT}_{\min}^2\right] - 2[(x - \text{PF}) \times E + \text{EnC}] \right. \\ \quad \left. \times T^2 \left(\dfrac{1}{x - y}\right)^3 \right\} \\ - (P - \text{EnC} \times \text{Fuel}) \times a \times T^2 \left(\dfrac{1}{x - y}\right)^3, \text{if } \dfrac{T}{\text{BT}_{\max}} + y < x \\ \quad \leq \dfrac{T}{\text{BT}_{\min}} + y \end{cases}$$

$$(14)$$

In Eq. (14) $x = \text{Power } P_1$, $y = \text{Power } P_2$, $a = \frac{S}{\text{BT}_{\max} - \text{BT}_{\min}}$.

We can see that when $y - \frac{T}{\text{BT}_{\max}} \leq x \leq y + \frac{T}{\text{BT}_{\max}}$, $\frac{\partial \text{Pro}_1}{\partial x} > 0$, so when the difference between x and y are not large enough to make any ship to change their choice of port both the ports tend to set the price just below $\frac{\text{Bunker}}{E}$. This is the highest price that the ships will choose to use shore power. The solution of partial derivative functions when $y - \frac{T}{\text{BT}_{\max}} > x \geq y - \frac{T}{\text{BT}_{\min}}$ or $\frac{T}{\text{BT}_{\max}} + y < x \leq \frac{T}{\text{BT}_{\min}} + y$ depends on the real value of the parameters such as P, EnC, $\text{BT}_{\max}$, $\text{BT}_{\min}$, PF. In another word, whether the port will lower its price for shore power depends on the profit that a ship with shore power facility will bring to the port authority. Since the profit function is not straightforward and kind of complex, the Nash equilibrium of the game should be discussed in various situations. Due to the capacity insufficiency, the thorough discussion will remain to be solved later.

4   Conclusion

This paper uses the Bertrand model to describe the price competition between two ports in a port group. The model is established successfully but due to the environmental cost the profit function becomes complicated and make it difficult to find the Nash equilibrium of the game. But the paper is still a start of investigation on the topic

of shore power pricing competition. Actually, the problem can be generalized to a pricing problem of upgraded product or service. Two companies decide to upgrade their product or service which has a higher marginal profit than the original ones. They try to identify the best price for an upgraded product to gain most profit. Customers have brand loyalty for the brand they originally choose, only when the other supplier provide a much better price will they change their mind. For the companies, a high price will lead to higher marginal profit but may lose some old customers. On the other hand, a lower price may attract customers from competitors but the marginal profit will be low. This problem can be observed in the market that the new generation of product keeps coming up. From personal opinion, this is an interesting topic that can be thoroughly studied.

References

1. Ballini, F., Bozzo, R.: Air pollution from ships in ports: the socioeconomic benefit of cold-ironing technology. Res. Transp. Bus. Manag. **17**, 92–98 (2015)
2. Chen, J., Zheng, T., Garg, A., Garg, A., Xu, L., Li, S., Fei, Y.: Alternative maritime power application as a green port strategy: barriers in China. J. Clean. Prod. **213**, 825–837 (2019)
3. Innes, A., Monios, J.: Identifying the unique challenges of installing cold ironing at small and medium ports: the case of aberdeen. Transp. Res. Part D **62**, 298–313 (2018)
4. Qu, X., Meng, Q.: The economic importance of the Straits of Malacca and Singapore: an extreme-scenario analysis. Transp. Res. Part E: Logistics Transp. Rev. **48**(1), 258–265 (2012)
5. Qu, X., Yu, Y., Zhou, M., Lin, C.T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach. Appl. Energy **257**, 114030 (2020)
6. Wang, S., Yan, R., Qu, X.: Development of a non-parametric classifier: effective identification, algorithm, and applications in port state control for maritime transportation. Transp. Res. Part B: Methodol. **128**, 129–157 (2019)
7. Transport & Environment comments to NSW EPA consultation regarding stricter sulphur fuel requirement for cruise ships in Sydney harbor (2015). Accessed 10th June, 2017. https://www.transportenvironment.org/
8. UNCTAD: Review of maritime transport. United Nations Publication, New York (2017)
9. Vaishnav, P., Fischbeck, P.S., Morgan, M.G., Corbett, J.J.: Shore power for vessels calling at US ports: benefits and costs. Environ. Sci. Technol. **50**(3), 1102–1110 (2016)
10. Winkel, R., Weddige, U., Johnsen, D., Hoen, V., Papaefthimiou, S.: Shore side electricity in Europe: potential and environmental benefits. Energy Policy **88**, 584–593 (2016)

Emission Evaluation of Marine Traffic

Jingwen Qi , **Shuaian Wang** , **and Xiaobo Qu**

Abstract Air pollution is an issue that has been widespread concern in all sectors of society. The pollutants, including toxic gases, greenhouse gases, and particulate matters, have permeated every aspect of our daily life and have a negative impact on human health, agriculture, industry, and climate change. Found by hard and thorough search, the human activities account for the majority and the transport sector is one of the most challenging areas, when it comes to abatement of local air pollution. Marine traffic, which covers over 80% of international trade, is mainly powered by cheap fuel oil with high impurities, so it will affect the social welfare of the coastal areas. Various measures that can be adopted to alleviate the problem to customize suitable regulations through research of the emission from shipping should be conducted. Also, the emission evaluation is critical to measure the efficiency of the regulation. Therefore, following the main steps of Ship Traffic Emissions Assessment Model, we summarize an activity-based framework of shipping emission evaluation that takes advantage of data from automatic identification system.

Keywords Emission evaluation · Activity-based method · AIS data

1 Introduction

Atmospheric pollution has always attracted the attention of international community. The main pollutants include sulphur oxides, carbon hydride, carbon oxides, nitrogen oxides, and particulate. These pollutants have adverse influence on many aspects

J. Qi (✉) · S. Wang
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong 999077, China
e-mail: jingwen.qi@connect.polyu.hk

S. Wang
e-mail: hans.wang@connect.polyu.hk

X. Qu
Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg 412 96, Sweden
e-mail: xiaobo@chalmers.se

such as human health, agriculture, industry, and climate change. The transport sector is one of the most challenging areas, when it comes to abatement of local air pollution. Marine traffic, which covers over 80% of the cargo volume of international trade [13], has become an important contributing factor because the large capacity vessels tend to use cheap fuel oil with high level of impurities. According to the greenhouse gas studies conducted by the international maritime organization in 2009 and 2014, exhaust emissions from shipping industry make up 15% nitrogen oxides, 13% Sulphur dioxide, and 2.7% carbon dioxide of global anthropogenic emission. Statistics shows that near 70% of the maritime emissions occur near port areas. And the coastal cities are always densely populated and highly developed. So, although the proportions are not astonishing, the threats of maritime emissions are not to be neglected. Different measures have been taken to reduce the marine traffic emission and mitigate the bad influence it may bring. The measures in different areas should be customized according to the situation and the efficiency of the measures needs to be calculated, so the work of emission evaluation is of great importance. In this paper, we summarize a framework of marine traffic exhaust emission evaluation.

For exhaust emission evaluation, the common methods can be divided into two types, fuel-based and movement-based approaches. A fuel-based approach evaluates exhaust emission of ships mainly with the fuel oil consumption data and makes little use of ship activities and movement information. When the fuel consumption details are available, namely the volume of fuel oil that has been consumed is already known, the emission volume can be easily calculated by multiplying the fuel usage and emission factor (emission volume exhausted by consuming unit mass fuel). Because of the simplicity of the input data and calculating process, this type of method requires computer hardware of lower function, but the results are rather rough. Therefore, fuel-based methods are used in few researches to estimate emissions for different countries and regions [4, 5, 9].

However, detailed fuel consumption data are difficult to obtain since the information is private to shipping companies, in other words, it is not known how much fuel oil has been used in a certain period. In fact, shipping companies are more interested in the average daily fuel usage or the mass of fuel oil consumed in a whole sailing route. This is not accurate enough to depict the distribution of marine traffic emissions. Instead, the more accessible location data with shorter time intervals can be used to calculate the consumed fuel amount, which is necessary for the exhaust emissions determination. So the other type of method, movement-based method, which takes advantage of ship movement data and information about the ship and its engine, is adopted in this paper. A movement-based approach firstly estimates the real-time fuel oil consumption rate according to the sailing speed inferred from the location data and technical details about the ship. Then, the emission volume is calculated on the basis of fuel oil amount that has been used. In practical terms, ships keep moving and the sailing speed is varying, as a result, the fuel oil consumption rate changes with time. So, compared with the fuel-based method, movement-based method can yield more accurate results in a short period. On the whole, a movement-based method requires a wide collection of data but is able to yield more precise results, and at the same time, match the emission at a certain timing node with

the location where it was exhausted. With the development of computing power, movement-based methods are extensively adopted by research papers as well as technical reports [2, 7, 10–12, 15, 17].

Existing researches about shipping emission evaluation can be divided into two types, academic papers and technical reports. Academic papers tend to put strength on the evaluation methods [4, 5, 7, 9, 10, 15]; they focus on the approach but usually make rough assumptions about the parameters. As a result, evaluation results of high precision requirements cannot be obtained by simply applying methods from papers. Meanwhile, technical reports [1, 11, 12, 16, 18], which are always conducted as a commission from government organizations or supported by them, have access to abundant and detailed data for the evaluation work. However, the approaches they use are more straightforward. At the same time, due to the various main purposes and databases, data of reports differs in level of detail and structure. Therefore, the data cannot be borrowed directly to use in a research. In order to find a scheme, that is, both operative and comparatively precise, this paper combines the method with data; reconciles them and summarizes an activity-based framework of shipping emission evaluation that takes advantage of data from automatic identification system and Ship Traffic Emissions Assessment Model.

2 Framework Description and Data Sources

The emission evaluation follows the main steps of Ship Traffic Emissions Assessment Model (STEAM), meanwhile, some parameters and technical details are obtained from a number of authoritative reports, such as those conducted by the IMO [1, 16] (and two reports led by Ng et al. [11, 12] with the help of the Marine Department and the Environmental Protection Department of the Hong Kong Special Administrative Region Government. STEAM is originally proposed by Jalkanen et al. [6] which has been used in a lot of emission evaluation works [7, 8, 10, 16] (Sofiev et al. [17]) after that. However, different from the original model, this paper does not take the performance penalty due to waves into consideration because of the absence of detailed historical data about sea waves. Also, from another aspect, the influence of waves for ships sailing in different areas along different directions can offset each other.

The input data of the calculation includes two types of information, properties of ships, as shown in Table 1, and vessel movement data.

The sailing trajectory and speed of a ship can be deduced from its coordinate at different times. For simplicity, transient speed is substituted with the average speed of a short period and the ship is assumed to sail in straight-line during the period. The shorter the data time interval, the more precise the trajectory and speed. Coordinates are sorted into hours to balance between the accuracy and the data processing time.

Table 1 Input data regarding ships' property

Property types	Description
Physical properties	Ship type
	Design speed v_{design} (knot)
Engine properties	Engine speed of main and auxiliary engines
	Total installed power of main engine $p_{installed}^{M}$ (kw)
Fuel properties	Fuel type used by the main and auxiliary engines Fuel T_M, Fuel T_A
	Sulphur content of different fuels SC_λ (mass%), $\lambda \in$ {HFO, MGO/MDO}

HFO: Heavy fuel oil; *MDO:* Marine diesel oil; *MGO:* Marine gas oil

3 Evaluation Steps

Based on location and time information extracted from the automatic identification system (AIS) and technical information about ships, the mass of SO_x, NO_x, CO_2, and PM emitted by ships can be positioned with a high spatial resolution. Before calculating emissions, the marine area is split by a grid, and each square denotes a small area.

We calculate the exhaust emission on a horizon which is equally divided into multiple time periods. Each period is defined by two timing nodes, e.g. the ith period is defined by the ith and $i + 1$th timing node. For each timing node, ship location information is given in the form of coordinate (x^i, y^i) more concretely x^i and y^i are the longitude and latitude of the ship's location at the ith timing node, respectively. In the evaluation process of this paper, east longitude and north latitude are set to be positive, naturally, the west longitude and south latitude are set to be negative. Given the coordinate data and technical parameters of the ship, the main steps to figure a ship's emission of NO_x, SO_x, CO_2, PM in the ith period are as follows.

Step 1: Obtain the ship's coordinate data at the ith and $i + 1$th timing node (x^i, y^i), (x^{i+1}, y^{i+1}).

Step 2: Obtain the sailing distance d^i (knot) and calculate the sailing speed v^i (knot) over the ith period, using the equations:

$$d^i = R \times \cos^{-1}\left[\cos y^i \times \cos y^{i+1} \cos(x^i - x^{i+1}) + \sin y^i \times \sin y^{i+1}\right] \div 1.852 \tag{1}$$

$$v^i = d^i \div T_{interval} \tag{2}$$

where R(km) is the radius of the earth and $T_{interval}$ is the span of the period in hours. Equation (1) is the spherical distance formula between two nodes, the constant 1.852 in Eq. (2) is used to convert the distance into nautical miles. Decide the ship operating mode according to v^i (knot).

$$\text{Operating mode} = \begin{cases} \text{cruise,} & v^i > 12 \\ \text{slow cruise,} & 12 \geq v^i > 8 \\ \text{maneuvering,} & 8 \geq v^i > 1 \\ \text{hotelling,} & 1 \geq v^i \end{cases} \tag{3}$$

Step 3: Calculate the transient power of the main engine p_M^i(kw), which can be evaluated as a function of v^i

$$p_M^i = \left(\frac{v^i}{v_{\text{design}}} \right)^3 \times \varepsilon_p \times p_{\text{installed}}^M \tag{4}$$

where v_{design} is the design speed, and the coefficient ε_p is assumed to be equal to 0.8 since the maximum power of the main engine is normally 80% of the installed power.

Step 4: Determine the transient power of the auxiliary engine. Information of auxiliary engine's power is not provided, so the data for different vessel types under each operating mode provided by the report conducted by Starcrest Consulting Group [18] for the Port of Los Angeles are put to use. For cruise, tanker, and container ships, the value is related to vessel capacity as shown in Tables 2, 3, and 4. Considering the technical advancement and improvement of mechanical efficiency on board, the auxiliary engine load is not strictly proportional to the vessel capacity, especially for

Table 2 Auxiliary engine load (p_A^i) for container ships (kw)

Capacity (TEU)	Operating mode		
	Cruise/slow cruise	Manoeuvring	Hotelling
0–1999	1122	2462	1396
2000–2999	766	2391	936
3000–3999	1629	2897	1638
4000–4999	2058	3766	1524
5000–5999	1635	2764	1605
6000–6999	1366	3556	3079
7000–7999	1722	3259	1570
8000–8999	1882	3555	1714
9000–9999	2684	2808	2031
10,000–10,999	2830	4075	2290
11,000–11,999	2790	3875	2570
12,000–12,999	2034	3002	1808
more than 13,000	1632	2719	1520

TEU: Twenty-foot equivalent unit

Table 3 Auxiliary engine load (p_A^i) for cruise ships (kw)

Passenger range	Operating mode		
	Cruise/slow cruise	Manoeuvring	Hotelling
0–1499	4404	5678	3479
1500–1999	7869	9869	6500
2000–2499	11,869	12,219	7769
2500–2999	10,650	9259	6958
3000–3499	9292	11,369	9292
more than 3499	10,945	12,411	11,445

Table 4 Auxiliary engine load (p_A^i) for tankers (kw)

DWT (ton)	Operating mode		
	Cruise/slow cruise	Manoeuvring	Hotelling
0–49,999	681	745	3406
50,000–120,000	728	1114	4044
more than 120,000	1004	1479	8992

DTW: Dead weight tonnage

Table 5 Auxiliary engine load (p_A^i) of four types of vessel (kw)

Vessel type	Operating mode		
	Cruise/slow cruise	Manoeuvring	Hotelling
Bulk	290	769	275
Reefer	617	1777	1194
RoRo	501	1449	1010
Miscellaneous	676	662	324

vary large container ships that are built recently. Power of auxiliary boiler has been included (Table 5).

Step 5: Estimate the NO_x emission $E_{NO_x}^i(g)$ of the main and auxiliary engine:

$$E_{NO_x}^i = T_{interval} \times \left[EF_{NO_x}^M \times p_M^i + EF_{NO_x}^A \times p_A^i \right] \tag{5}$$

NO_x emission factor for main engine $EF_{NO_x}^M$ is related to ship operating mode and engine speed while $EF_{NO_x}^A$ emission factor for auxiliary engine, only depends on engine speed as shown in Table 6 referring to European Commission [3] and Ng et al. [12]. In the table, SSD, MSD and HSD represent the slow speed diesel, medium speed diesel, and high speed diesel, respectively.

Step 6: Determine the fuel oil consumption rate of the main engine $FOCR_M$. Main engines are used for propulsion and $FOCR_M$ is related to the transient sailing

Table 6 Emission factors for main/auxiliary engine (gNO$_x$/kwh)

Engine speed	Main engine		Auxiliary engine
	Hotelling/manoeuvring	Cruise/slow cruise	
SSD	13.6	17.0	NA
MSD	10.6	13.2	13.9
HSD	9.6	12.0	13.9

Table 7 Fuel oil consumption rate baseline of main engine (g fuel/kwh)

Engine year of build	SSD	MSD	HSD
Before 1993	205	215	225
1994–2010	185	195	205
Post 2011	175	185	195

speed and output power, therefore, the fuel oil consumption rate baseline of the main engine $FOCR_M^B$(g fuel/kwh) is introduced. Firstly determine $FOCR_M^B$ (g fuel/kwh) according to Table 7 referring to IMO (2009). As is shown in Table 4, $FOCR_M^B$ is closely related to the engine speed and the year of construction, old engines with high speed tend to consume more fuel oil to output a power unit.

Then calculate the transient value of fuel oil consumption rate. Following IMO (2014), the equation:

$$FOCR_M = FOCR_M^B \times \left(0.455 \times \left(\frac{p^i}{p_{installed}^M} \right)^2 - 0.71 \times \frac{p^i}{p_{installed}^M} + 1.28 \right) \quad (6)$$

is adopted to explain the relationship between $FOCR_M$ and $FOCR_M^B$ under different output power. As revealed by Eq. 6, main engine is most efficient at around 80% load, with either higher or lower load, and it will take more oil for the engine to do the same work.

Step 7: Determine the fuel oil consumption rate of auxiliary engine $FOCR_A$. As stated by Cooper and Gustafsson [2], $FOCR_A$ depends on the fuel type that the auxiliary engine consumes. 227 grams of oil is needed to output 1 kilowatt-hour for auxiliary engines with heavy fuel oil (HFO), for those consume marine diesel oil (MDO)/marine gas oil (MGO) $FOCR_A$ equal to 217 (g fuel/kwh).

Step 8: Estimate the emission of SO$_x$ according to the following equation:

$$E_{SO_x}^i = T_{interval} \times \left[FOCR_M \times p_M^i \times SC_{Fuel\,T_A} + FOCR_A \times p_A^i \times SC_{Fuel\,T_A} \right] \times \frac{M_{SO_2}}{M_S} \quad (7)$$

where $M_{SO_2} = 64$, $M_S = 32$ are the molar mass (g/mol) of sulphur dioxide and sulphur, respectively, $SC_{Fuel\,T_M}$ and $SC_{Fuel\,T_A}$ are sulphur content of fuel for main and

Table 8 CO_2 emission factor of different fuel types (g CO_2/g fuel)

Type of fuel λ	Emission factor $EF_{CO_2,\lambda}$
MDO/ MGO	3.206
HFO	3.114

auxiliary engine. It is assumed that all the sulphur element is transferred into SO_2 through the process of combustion.

Step 9: Estimate the emission of CO_2. It is assumed that the emission factories only affected by fuel type. In this study, the emission factor data of different fuel types come from MEPC 63/23, annex 8 (Table 8).

The CO_2 emission can be evaluated as:

$$E_{CO_2}^i = T_{interval} \times \left[FOCR_M \times p_M^i \times EF_{CO_2,FuelT_M} + FOCR_A \times p_A^i \times EF_{CO_2,FuelT_A} \right]. \tag{8}$$

Step 10: Calculate the emission of particulate matter (PM) E_{PM}^i.

$$E_{PM}^i = E_{PM,M}^i + E_{PM,A}^i \tag{9}$$

In Eq. (9), $E_{PM,M}^i$ and $E_{PM,A}^i$ are the PM emission from the main and auxiliary engine, respectively. According to IMO (2014), the emission volume of PM is related to fuel type, so different equations are applied to evaluate the PM exhausted emission for engines that consume different fuel oil:

$$E_{PM,\mu}^i = \begin{cases} T_{interval} \times p_\mu^i \left[1.35 + FOCR_\mu \times 0.157 \times \left(SC_{FuelT_\mu} - 0.0246 \right) \right], \\ \text{Fuel } T_\mu = HFO \\ T_{interval} \times p_\mu^i \left[0.23 + FOCR_\mu \times 0.157 \times \left(SC_{FuelT_\mu} - 0.0024 \right) \right], \\ \text{Fuel } T_\mu = MDO/MGO. \end{cases} \tag{10}$$

In the above equation, $\mu \in \{M, A\}$ represents the engine type.

Step 11: Add $E_{NO_x}^i$, $E_{SO_x}^i$, $E_{CO_2}^i$, E_{PM}^i to the total emissions of the square in which the midpoint of the trajectory $(\frac{x^i+x^{i+1}}{2}, \frac{y^i+y^{i+1}}{2})$ locates.

Repeat the above 11 steps until all the data for the evaluation horizon are processed, and then the distribution of emissions of a ship is completed. The whole picture of marine traffic exhaust emissions is drawn by combining the results of all the ships.

4 Assumptions About Fuel Oil Type and Sulphur Content

Fuel type and fuel sulphur content are two important variables in the exhaust emission evaluation. In practice, to lower the bunker cost, HFO is always used for propulsion

in large ships, at the same time, a number of ships with lower main engine power may use MDO/MGO. For auxiliary engines, it is common to use MDO/MGO. However, data about the fuel type chosen by each ship is not provided, so corresponding assumptions are made in academic papers and technical reports.

In papers adopting fuel-based method fuel type is known. In researches using activity-based method certain assumptions are made. A number of papers focus on proposing a new method so, for simplicity, they tend to assume certain sulphur content for all the fuel oil [6]. Others [8, 10, 17] always assume that fuel oil with the highest sulphur content allowed in the area according to relative regulations is used. However, technical reports conducted by organizations that have access to more complete information make more precise assumptions on fuel type and fuel quality. In Ng et al. [11], it is assumed that vessels with main engine power larger than 1100 kw burn HFO, and those with lower main engine power burn MDO/MGO. Meanwhile, sulphur contents of HFO used by main engine auxiliary engine and auxiliary boiler are assumed to be 2.83%, 2.64%, and 2.77%, respectively, and MDO/MGO are assumed to contain 0.5% sulphur in mass. More concretely, IMO [1] lists the common fuel type used by ships with various types and sizes. IMO [16] summarizes the average sulphur content for both HFO and MDO/MGO fuels from 2007 to 2012.

In IMO [1], the fuel types for various ships are not listed explicitly, for some ships of a certain size both HFO and MDO/MGO can be used according to the report. For emission evaluation adopting bottom-up approach, better defined assumptions are required, so in this paper, assumptions are made following Ng et al. [11] while statistical data from IMO [16] are considered. For fuel types, auxiliary engines consume MDO/MGO regardless of ship type and size. At the same time, main engines with installed power higher than 3000 kw consume HFO, and those with lower installed power consume MDO/MGO. According to IMO, the latest figures showed that the average sulphur content of HFO tested in 2017 is 2.54%. The worldwide average sulphur content for MDO/MGO in 2017 is 0.08%. Additionally, when sailing in the emission control area (ECA), ships will adopt measures to obey the restriction on fuel oil sulphur content, e.g. switching fuel oil or installing sulphur scrubbers. We expect that machine learning-based approaches [14, 19] will be of value to predict the emissions from ships more effectively.

5   Conclusion

We have summarized a marine traffic emission evaluation framework based on the literature. The framework can be used by practitioners and researchers to calculate the emission in a particular area, which is indispensable of studies in the area of emission control and green shipping. We hope that in future there will be more research on these topics and make the shipping industry more sustainable and environmentally friendly.

References

1. Buhaug, Ø., Corbett, J.J., Endresen, Ø., Eyring, V., Faber, J., Hanayama, S., Lee, D.S., Lee, D., Lindstad, H., Markowska, A.Z., Mjelde, A., Nelissen, D., Nilsen, J., Pålsson, C., Winebrake, J.J., Wu, W., Yoshida, K.: Second IMO GHG Study. International Maritime Organization (IMO), London, UK, April (2009)
2. Cooper D., Gustafsson T.: Methodology for calculating emissions from ships: 1. Update of emission factors. Swedish Environmental Protection Agency (2004, February)
3. European Commission: Quantification of emissions from ships associated with ship movements between ports in the European Community. Entec UK Ltd and IVL Swedish Environmental Research Institute (2002, July)
4. Endresen, O., Bakke, J., Sorgard, E., Berglen, T.F., Holmvang, P.: Improved modelling of ship SO_2 emissions—a fuel-based approach. Atmos. Environ. **39**, 3621–3628 (2005)
5. Hulskotte, J.H.J., Denier, H.A.C.: Fuel consumption and associated emissions from seagoing ships at berth derived from an on-board survey. Atmos. Environ. **44**, 1229–1236 (2010)
6. Jalkanen, J.P., Brink, A., Kalli, J., Pettersson, H., Kukkonen, J., Stipa, T.: A modelling system for the exhaust emissions of marine traffic and its application in the Baltic Sea area. Atmos. Chem. Phys. **9**(23), 9209–9223 (2009)
7. Johansson, L., Jalkanen, J.P., Kalli, J., Kukkonen, J.: The evolution of shipping emissions and the costs of regulation changes in the northern EU area. Atmos. Chem. Phys. **13**, 11375–11389 (2013)
8. Jonson, J., Jalkanen, J., Johansson, L., Gauss, M., Denier, H.A.C.: Model calculations of the effects of present and future emissions of air pollutants from shipping in the Baltic Sea and the North Sea. Atmos. Chem. Phys. **15**, 783–798 (2015)
9. Kesgin, U., Vardar, N.: A study on exhaust gas emissions from ships in Turkish Straits. Atmos. Environ. **35**, 1863–1870 (2001)
10. Marelle, L., Thomas, J.L., Raut, J.C., Law, K.S., Jalkanen, J.P., Johansson, L., Roiger, A., Schlager, H., Kim, J., Reiter, A., Weinzierl, B.: Air quality and radiative impacts of Arctic shipping emissions in the summertime in northern Norway: From the local to the regional scale. Atmos. Chem. Phys. **16**, 2359–2379 (2016)
11. Ng, K.W.S., Lin, C., Chan, J.W.M., Yip, A.C.K., Lau, A.K.H., Fung, J.C.: Study on Marine Vessels Emission Inventory: Final Report. Environmental Protection Department. The HKSAR Government. http://www.epd.gov.hk/epd/sites/default/files/epd/english/environmentinhk/air/studyrpts/files/EPD_MVEIS_Final_Report_v7_approved_120528.pdf (2012)
12. Ng, K.W.S., Zheng, A.J., Li, C., Ou, J., Fan, X.: Pearl River Delta Ship Emission Inventory Study TECHNICAL REPORT EN. https://civic-exchange.org/report/pearl-river-delta-ship-emission-inventory-study/ (2016)
13. Qu, X., Meng, Q.: The economic importance of the Straits of Malacca and Singapore: an extreme-scenario analysis. Transp. Res. Part E: Logistics Transp. Rev. **48**(1), 258–265 (2012)
14. Qu, X., Yu, Y., Zhou, M., Lin, C.T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach. Appl. Energy **257**, 114030 (2020)
15. Schrooten, L., Vlieger, I.D., Panis, L.I., Styns, K., Torfs, R.: Inventory and forecasting of maritime emissions in the Belgian sea territory, an activity-based emission model. Atmos. Environ. **42**(4), 667–676 (2008)
16. Smith, T.W.P., Jalkanen, J.P., Anderson, B.A., Corbett, J.J., Faber, J., Hanayama, S., Aldous, L., et al.: Third IMO GHG Study. International Maritime Organization, London, UK (2014)

17. Sofiev, M., Winebrake, J.J., Johansson, L., Carr, E.W., Prank, M., Soares, J., Vira, J., Kouznetsov, R., Jalkanen, J.P., Corbett, J.J.: Cleaner fuels for ships provide public health benefits with climate tradeoffs. Nat. Commun. **9**(1), 406 (2018)
18. Starcrest Consulting Group Port of Los Angeles Inventory of Air Emissions 2009. Technical Report (2018, July)
19. Wang, S., Yan, R., Qu, X.: Development of a non-parametric classifier: effective identification, algorithm, and applications in port state control for maritime transportation. Transp. Res. Part B: Methodol. **128**, 129–157 (2019)

Correction to: Injury Severity Analysis of Secondary Incidents

Jing Li and Jingqiu Guo

Correction to:
Chapter "Injury Severity Analysis of Secondary Incidents"
in: X. Qu et al. (eds.),
***Smart Transportation Systems 2020*, Smart Innovation,**
Systems and Technologies 185,
https://doi.org/10.1007/978-981-15-5270-0_14

In the original version of the book, in Chapter 14, the following belated corrections in references were incorporated.

(1) added to acknowledge:

"This is a short initial version, for more detailed see Li et al. (2020)"

(2) added to reference: "Li, J., Guo, J., Wijnands, J., Yu, R., Xu, C., Stevenson, M.: Assessing injury severity of secondary incidents using support vector machines. Journal of Transportation Safety & Security, 1-20 (2020)"

The chapter and book have been updated with the changes.

Author Index